PATENT COOPERATION TREATY

From the INTERNATIONAL BUREAU

PCT

NOTIFICATION OF ELECTION

(PCT Rule 61.2)

To:

United States Patent and Trademark
Office
(Box PCT)
Crystal Plaza 2
Washington, DC 20231
ÉTATS-UNIS D'AMÉRIQUE

in its capacity as elected Office

Date of mailing (day/month/year) 29 April 1999 (29.04.99)	
International application No. PCT/GB98/02630	**Applicant's or agent's file reference** P22410/PKE/BOU
International filing date (day/month/year) 02 September 1998 (02.09.98)	**Priority date** (day/month/year) 02 September 1997 (02.09.97)
Applicant McGREGOR, Duncan	

1. The designated Office is hereby notified of its election made:

☒ in the demand filed with the International Preliminary Examining Authority on:

17 March 1999 (17.03.99)

☐ in a notice effecting later election filed with the International Bureau on:

2. The election ☒ was
☐ was not

made before the expiration of 19 months from the priority date or, where Rule 32 applies, within the time limit under Rule 32.2(b).

The International Bureau of WIPO **34, chemin des Colombettes** **1211 Geneva 20, Switzerland** Facsimile No.: (41-22) 740.14.35	Authorized officer Lazar Joseph Panakal Telephone No.: (41-22) 338.83.38

TENT COOPERATION TREAT

PCT

INTERNATIONAL SEARCH REPORT

(PCT Article 18 and Rules 43 and 44)

Applicant's or agent's file reference P22410/PKE/BOU	**FOR FURTHER ACTION** see Notification of Transmittal of International Search Report (Form PCT/ISA/220) as well as, where applicable, item 5 below.	
International application No. PCT/GB 98/ 02630	International filing date *(day/month/year)* 02/09/1998	(Earliest) Priority Date *(day/month/year)* 02/09/1997
Applicant ROWETT RESEARCH SERVICES LIMITED et al.		

This International Search Report has been prepared by this International Searching Authority and is transmitted to the applicant according to Article 18. A copy is being transmitted to the International Bureau.

This International Search Report consists of a total of 3 sheets.

☒ It is also accompanied by a copy of each prior art document cited in this report.

1. ☐ **Certain claims were found unsearchable** (see Box I).

2. ☐ **Unity of invention is lacking** (see Box II).

3. ☒ The international application contains disclosure of a **nucleotide and/or amino acid sequence listing** and the international search was carried out on the basis of the sequence listing
   - ☒ filed with the international application.
   - ☐ furnished by the applicant separately from the international application,
     - ☐ but not accompanied by a statement to the effect that it did not include matter going beyond the disclosure in the international application as filed.
   - ☐ Transcribed by this Authority

4. With regard to the **title**,
   - ☒ the text is approved as submitted by the applicant
   - ☐ the text has been established by this Authority to read as follows:

5. With regard to the **abstract**,
   - ☒ the text is approved as submitted by the applicant
   - ☐ the text has been established, according to Rule 38.2(b), by this Authority as it appears in Box III. The applicant may, within one month from the date of mailing of this International Search Report, submit comments to this Authority.

6. The figure of the **drawings** to be published with the abstract is:

   Figure No. –

   - ☐ as suggested by the applicant.
   - ☐ because the applicant failed to suggest a figure.
   - ☐ because this figure better characterizes the invention.

   ☐ None of the figures.

INTERNAT[illegible]NAL SEARCH REPORT

Ir. [illegible]onal Application No
PCT/GB 98/02630

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/12 C12N15/10 C12N15/62 C07K14/72 C07K16/00
C12Q1/68 G01N33/577

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N C07K C12Q G01N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 93 08278 A (AFFYMAX TECH NV) ✓ 29 April 1993 see page 9, line 32 - line 37 ---	1-23
A	US 5 597 693 A (EVANS RONALD M ET AL) 28 January 1997 see column 1, line 38 - line 49 see column 1, line 56 - line 61 ---	1-23
A	M. BROWN AND P.A. SHARP: "Human estrogen receptor forms multiple protein-DNA complexes" J. BIOL. CHEM., vol. 265, no. 19, 5 July 1990, pages 11238-11243, XP002092292 AM. SOC. BIOCHEM. MOL.BIOL.,INC.,BALTIMORE,US see the whole document --- -/--	1-23

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search 4 February 1999	Date of mailing of the international search report 18/02/1999
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Authorized officer Hornig, H

1

INTERNAT[illegible]NAL SEARCH REPORT

Ir. [illegible]al Application No
PCT/GB 98/02630

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	WO 92 20791 A (CMABRIDGE ANTIBODY TECH ;MEDICAL RES COUNCIL (GB)) ✓ 26 November 1992 cited in the application see the whole document ---	1-23
A	DE 196 46 372 C (EVOTEC BIOSYSTEMS GMBH) 19 June 1997 see the whole document ---	1-23
A	WO 92 01047 A (CAMBRIDGE ANTIBODY TECH ;MEDICAL RES COUNCIL (GB)) 23 January 1992 ✓ cited in the application see the whole document -----	1-23

1

INTERNAT[illegible]AL SEARCH REPORT

Information on patent family members

I[illegible]ial Application No
PCT/GB 98/02630

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9308278 A	29-04-1993	US 5270170 A	14-12-1993
		AU 3759693 A	21-05-1993
		EP 0610448 A	17-08-1994
		US 5498530 A	12-03-1996
		US 5733731 A	31-03-1998
		US 5338665 A	16-08-1994
US 5597693 A	28-01-1997	AT 166360 T	15-06-1998
		AU 655912 B	19-01-1995
		AU 5342390 A	22-10-1990
		CA 2047752 A	18-09-1990
		DE 69032324 D	25-06-1998
		DE 69032324 T	08-10-1998
		EP 0463081 A	02-01-1992
		JP 4505012 T	03-09-1992
		WO 9011273 A	04-10-1990
WO 9220791 A	26-11-1992	AT 145237 T	15-11-1996
		AU 664155 B	09-11-1995
		AU 8221691 A	04-02-1992
		CA 2086936 A	11-01-1992
		DE 69123156 D	19-12-1996
		DE 69123156 T	17-04-1997
		DK 589877 T	07-04-1997
		EP 0589877 A	06-04-1994
		EP 0585287 A	09-03-1994
		EP 0774511 A	21-05-1997
		EP 0844306 A	27-05-1998
		ES 2096655 T	16-03-1997
		WO 9201047 A	23-01-1992
		GR 3022126 T	31-03-1997
		AU 665190 B	21-12-1995
		AU 1693892 A	30-12-1992
		CA 2109602 A	26-11-1992
		JP 6508511 T	29-09-1994
		AU 665025 B	14-12-1995
		AU 2593392 A	27-04-1993
		AU 665221 B	21-12-1995
		AU 3089092 A	28-06-1993
		AU 673515 B	14-11-1996
		AU 3763893 A	21-10-1993
		CA 2119930 A	01-04-1993
		CA 2124460 A	10-06-1993
		CA 2131151 A	30-09-1994
		EP 0605522 A	13-07-1994
		EP 0616640 A	28-09-1994
		EP 0656941 A	14-06-1995
		WO 9306213 A	01-04-1993
		WO 9311236 A	10-06-1993
		WO 9319172 A	30-09-1993
		JP 6510671 T	01-12-1994
		JP 7502167 T	09-03-1995
		JP 7505055 T	08-06-1995
		US 5565332 A	15-10-1996
		US 5733743 A	31-03-1998
DE 19646372 C	19-06-1997	NONE	

INTERNAT AL SEARCH REPORT

Information on patent family members

I[illegible]onal Application No
PCT/GB 98/02630

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9201047 A	23-01-1992	AT 145237 T	15-11-1996
		AU 664155 B	09-11-1995
		AU 8221691 A	04-02-1992
		CA 2086936 A	11-01-1992
		DE 69123156 D	19-12-1996
		DE 69123156 T	17-04-1997
		DK 589877 T	07-04-1997
		EP 0589877 A	06-04-1994
		EP 0585287 A	09-03-1994
		EP 0774511 A	21-05-1997
		EP 0844306 A	27-05-1998
		ES 2096655 T	16-03-1997
		WO 9220791 A	26-11-1992
		GR 3022126 T	31-03-1997
		AU 665190 B	21-12-1995
		AU 1693892 A	30-12-1992
		CA 2109602 A	26-11-1992
		JP 6508511 T	29-09-1994

17

PATENT COOPERATION TREATY

REC'D 23 DEC 1999
WIPO PCT

PCT

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

(PCT Article 36 and Rule 70)

Applicant's or agent's file reference P22410/PKE/BOU	**FOR FURTHER ACTION**	See Notification of Transmittal of International Preliminary Examination Report (Form PCT/IPEA/416)
International application No. PCT/GB98/02630	International filing date *(day/month/year)* 02/09/1998	Priority date *(day/month/year)* 02/09/1997

International Patent Classification (IPC) or national classification and IPC
C12N15/12

Applicant
ROWETT RESEARCH SERVICES LIMITED et al.

1. This international preliminary examination report has been prepared by this International Preliminary Examining Authority and is transmitted to the applicant according to Article 36.

2. This REPORT consists of a total of 5 sheets, including this cover sheet.

   ☐ This report is also accompanied by ANNEXES, i.e. sheets of the description, claims and/or drawings which have been amended and are the basis for this report and/or sheets containing rectifications made before this Authority (see Rule 70.16 and Section 607 of the Administrative Instructions under the PCT).

   These annexes consist of a total of sheets.

3. This report contains indications relating to the following items:

   I ☒ Basis of the report
   II ☒ Priority
   III ☐ Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
   IV ☐ Lack of unity of invention
   V ☒ Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations suporting such statement
   VI ☐ Certain documents cited
   VII ☐ Certain defects in the international application
   VIII ☒ Certain observations on the international application

Date of submission of the demand 17/03/1999	Date of completion of this report 17.12.99
Name and mailing address of the international preliminary examining authority: European Patent Office D-80298 Munich Tel. +49 89 2399 - 0 Tx: 523656 epmu d Fax: +49 89 2399 - 4465	Authorized officer Herrmann, K Telephone No. +49 89 2399 2670

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No. PCT/GB98/02630

I. Basis of the report

1. This report has been drawn on the basis of (*substitute sheets which have been furnished to the receiving Office in response to an invitation under Article 14 are referred to in this report as "originally filed" and are not annexed to the report since they do not contain amendments.*):

Description, pages:

1-58	as originally filed

Claims, No.:

1-23	as originally filed

Drawings, sheets:

1/9-9/9	as originally filed

2. The amendments have resulted in the cancellation of:

☐ the description, pages:
☐ the claims, Nos.:
☐ the drawings, sheets:

3. ☐ This report has been established as if (some of) the amendments had not been made, since they have been considered to go beyond the disclosure as filed (Rule 70.2(c)):

4. Additional observations, if necessary:

see separate sheet

II. Priority

1. ☐ This report has been established as if no priority had been claimed due to the failure to furnish within the prescribed time limit the requested:

☐ copy of the earlier application whose priority has been claimed.

☐ translation of the earlier application whose priority has been claimed.

2. ☐ This report has been established as if no priority had been claimed due to the fact that the priority claim has

been found invalid.

Thus for the purposes of this report, the international filing date indicated above is considered to be the relevant date.

3. Additional observations, if necessary:

see separate sheet

V. Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

1. Statement

Novelty (N)	Yes:	Claims	1-22
	No:	Claims	23
Inventive step (IS)	Yes:	Claims	1-22
	No:	Claims	23
Industrial applicability (IA)	Yes:	Claims	1-23
	No:	Claims	

2. Citations and explanations

see separate sheet

VIII. Certain observations on the international application

The following observations on the clarity of the claims, description, and drawings or on the question whether the claims are fully supported by the description, are made:

see separate sheet

Citations

The documents mentioned in this International Preliminary Examination Report (IPER) are numbered as in the International Search Report dated 04.02.99, i.e. **D1** and **D6** correspond to the first and the last document of the search report, respectively.

Re ITEM I (Basis of the opinion)

1 Description pages 17, 25 and 26 filed with letter of 23.12.98 are regarded as originally filed.

2 The sequence listing consisting of 28 pages (78 sequences) filed with letter of 23.12.98 can be regarded as meeting the requirements of Art. 34(2)(b) PCT.

3 The application contains 3 deposit receipts (NCIMB 40970, 40971, 40972).

Re ITEM II (Priority)

The priority document pertaining to the present application was not available at the time this IPER was established. Hence, the current assessment is based on the assumption that all claims enjoy priority rights from the filing date of the priority document (02.09.97).

Re ITEM V (Novelty, inventive step, industrial applicability)

1 Novelty and inventive step (Art. 33(2) and (3) PCT)

1.1 The subject-matter of claims 1-22 has not been made available to the public by any of the available prior art documents and can therefore be regarded as novel.

1.2 Due to the open claim language recited in independent claim 23, the subject-matter of said claim does not meet the requirements of Art. 33(2) and 33(3) PCT.

The term "substantially" renders the scope of protection so broad that any prior art sequence which is homologous to the sequences as set out in SEQ ID NOs 15 and 17 deprives said claim of novelty (see e.g. **D4**, Fig. 7 and **D6**, Fig. 13).

2 Inventive step (Art. 33(3) PCT)

The subject-matter of claims 1-22 cannot be derived from the available prior art in an obvious manner and therefore complies with the requirements of Art. 33(3) PCT.

3 Industrial applicability (Art. 33(4) PCT)

Claims 1-23 meet the criteria as set forth by Art. 33(4) PCT.

Re ITEM VIII (Certain observations on the international application)

1 How the effect of protecting the "chimeric protein-encoding portion of the recombinant polynucleotide not bound by the chimeric protein nucleotide binding portion" is achieved is not explained, i.e. is not defined by any technical features in claim 1. According to claims 11 and 12, the recombinant polynucleotide encodes *inter alia* a non-sequence-specific nucleotide binding protein, namely a viral coat protein. This essential technical feature is however not included in claims 1, 20(f) and 21(f) (Art. 6 PCT).

2 The category of claim 10 is rendered unclear by the present wording of said claim (Art. 6 PCT). Claim 10 is a product claim directed to a peptide display carrier package (PDCP). Said PDCP is not defined in more detail in claim 10. Merely, methodological steps are given which do not define the PDCP *per se* in more detail.

Chimeric binding peptide library screening method

The present invention relates generally to methods for screening nucleotide libraries for sequences that encode peptides of interest.

Isolating an unknown gene which encodes a desired peptide from a recombinant DNA library can be a difficult task. The use of hybridisation probes may facilitate the process, but their use is generally dependent on knowing at least a portion of the sequence of the gene which encodes the protein. When the sequence is not known, DNA libraries can be expressed in an expression vector, and antibodies have been used to screen for plaques or colonies displaying the desired protein antigen. This procedure has been useful in screening small libraries, but rarely occurring sequences which are represented in less than about 1 in 10^5 clones (as is the case with rarely occurring cDNA molecules or synthetic peptides) can be easily missed, making screening libraries larger than 10^6 clones at best laborious and difficult. Methods designed to address the isolation of rarely occurring sequences by screening libraries of 10^6 clones have been developed and include phage display methods and LacI fusion phage display, discussed in more detail below.

Phage display methods. Members of DNA libraries which are fused to the N-terminal end of filamentous bacteriophage pIII and pVIII coat proteins have been expressed from an expression vector resulting in the display of foreign peptides on the surface of the phage particle with the DNA encoding the fusion protein packaged in the phage particle (Smith G. P., 1985, Science 228: 1315-1317). The expression vector can be the bacteriophage genome itself, or a phagemid vector, into which a bacteriophage coat protein has been cloned. In the latter case, the host bacterium, containing the phagemid vector, must be co-infected with autonomously replicating bacteriophage, termed helper phage, to provide the full complement of proteins necessary to produce mature phage particles. The helper phage normally has a genetic defect in the origin of replication which results in the preferential packaging of the phagemid genome. Expression of the fusion protein following helper phage infection, allows incorporation of both fusion protein and wild type coat protein into the phage particle during assembly. Libraries of fusion proteins incorporated into phage, can then be selected for binding members against targets of interest (ligands). Bound phage can then be allowed to reinfect *Escherichia coli* (*E. coli*) bacteria and then amplified and the selection repeated, resulting in the enrichment of binding members (Parmley, S. F., & Smith, G. P. 1988., Gene 73: 305-318; Barrett R. W. *et al*., 1992, Analytical Biochemistry 204: 357-364 Williamson *et al*., Proc. Natl. Acad. Sci. USA, 90: 4141-4145; Marks *et al*., 1991, J. Mol. Biol. 222: 581-597).

Several publications describe this method. For example, US Patent No 5,403,484 describes production of a chimeric protein formed from the viral coat protein and the peptide of interest. In this method at least a functional portion of a viral coat protein is required to cause display of the chimeric protein or a processed form thereof on the outer surface of the virus. In addition, US Patent No 5,571,698 describes a method for obtaining a nucleic acid encoding a binding protein, a key component of which comprises preparing a population of amplifiable genetic packages which have a genetically determined outer surface protein, to cause the display of the potential binding domain on the outer surface of the genetic package. The genetic packages are selected from the group consisting of cells, spores and viruses. For example when the genetic package is a bacterial cell, the outer surface transport signal is derived from a bacterial outer surface protein, and when the genetic package is a filamentous bacteriophage, the outer surface transport signal is provided by the gene pIII (minor coat protein) or pVIII (major coat protein) of the filamentous phage.

WO-A-92/01047 and WO-A-92/20791 describe methods for producing multimeric specific binding pairs, by expressing a first polypeptide chain fused to a viral coat protein, such as the gene pIII protein, of a secreted replicable genetic display package (RGDP) which displays a polypeptide at the surface of the package, and expressing a second polypeptide chain of the multimer, and allowing the two chains to come together as part of the RGDP.

LacI fusion plasmid display. This method is based on the DNA binding ability of the lac repressor. Libraries of random peptides are fused to the lacI repressor protein, normally to the C-terminal end, through expression from a plasmid vector carrying the fusion gene. Linkage of the LacI-peptide fusion to its encoding DNA occurs via the lacO sequences on the plasmid, forming a stable peptide-LacI-peptide complex. These complexes are released from their host bacteria by cell lysis, and peptides of interest isolated by affinity purification on an immobilised target. The plasmids thus isolated can then be reintroduced into *E. coli* by electroporation to amplify the selected population for additional rounds of screening (Cull, M. G. et al. 1992. Proc. Natl. Acad. Sci. U.S.A. 89:1865-1869).

US Patent No 5498530 describes a method for constructing a library of random peptides fused to a DNA binding protein in appropriate host cells and culturing the host cells under conditions suitable for expression of the fusion proteins intra-cellularly, in the cytoplasm of the host cells. This method also teaches that the random peptide is located at the carboxy terminus of the fusion protein and that the fusion protein-DNA complex is released from the host cell by cell lysis. No method is described for the protection of the DNA from degradation once released from the lysed cell. Several DNA binding proteins are claimed but no examples are shown except lacI.

There remains a need for methods of constructing peptide libraries in addition to the methods described above. For instance, the above methods do not permit production of secreted peptides with a free carboxy terminus. The present invention describes an alternative method for isolating peptides of interest from libraries and has significant advantages over the prior art methods.

utility?

In general terms, the present invention provides a method for screening a nucleotide library (usually a DNA library) for a nucleotide sequence which encodes a target peptide of interest. The method involves physically linking each peptide to a polynucleotide including the specific nucleotide sequence encoding that peptide. Linkage of a peptide to its encoding nucleotide sequence is achieved via linkage of the peptide to a nucleotide binding domain. A bifunctional chimeric protein with a nucleotide binding domain and a library member or target peptide (preferably with a function of interest) is thus obtained. The peptide of interest is bound to the polynucleotide encoding that peptide via the nucleotide binding domain of the chimeric protein.

The polynucleotide-chimeric protein complex is then incorporated within a peptide display carrier package (PDCP), protecting the polynucleotide from subsequent degradation, while displaying the target peptide portion on the outer surface of the peptide display carrier package (PDCP).

Thus, in one aspect, the present invention provides a peptide display carrier package (PDCP), said package comprising a polynucleotide-chimeric protein complex wherein the chimeric protein has a nucleotide binding portion and a target peptide portion, wherein said polynucleotide comprises a nucleotide sequence motif which is specifically bound by said nucleotide binding portion, and wherein at least the chimeric protein encoding portion of the polynucleotide not bound by the nucleotide binding portion of the chimeric protein is protected.

In one embodiment the polynucleotide is protected by a protein which binds non-specifically to naked polynucleotide. Examples include viral coat proteins, many of which are well-known in the art. Where the chosen viral coat protein requires an initiation sequence to commence general binding to the polynucleotide, this will be provided on the polynucleotide at appropriate location(s). A preferred coat protein is coat protein from a bacteriophage, especially M13.

Generally, the nucleic binding portion of the chimeric protein is selected for its specificity for the nucleotide sequence motif present in the recombinant polynucleotide encoding the chimeric protein.

Optionally, the nucleotide sequence motif may be an integral part of the protein encoding region of the polynucleotide. Alternatively, and more usually, the motif may be present in a non-coding region of the polynucleotide. For the purposes of this invention, all that is required is for the motif to be located on the polynucleotide such that the nucleotide binding portion of the chimeric protein is able to recognise and bind to it. Desirably the polynucleotide-chimeric protein complex has a dissociation constant of at least one hour.

Optionally, the recombinant polynucleotide may comprise two or more nucleotide sequence motifs, each of which will be bound by a chimeric protein molecule. Preferably, the motifs are positioned along the length of the polynucleotide to avoid steric hindrance between the bound chimeric proteins.

Preferably, the nucleotide sequence motif is not affected by the presence of additional nucleotide sequence (e.g. encoding sequence) at its 5′ and/or 3′ ends. Thus the chimeric fusion protein may include a target peptide portion at its N terminal end, at its C terminal end or may include two target peptide portions (which may be the same or different) at each end of the nucleotide binding portion, ie at both the N and C terminal ends of the chimeric protein. For example one target peptide may be an antibody of known specificity and the other peptide may be a peptide of potential interest.

Desirably the target peptide portion of the chimeric protein is displayed externally on the peptide display carrier package, and is thus available for detection, reaction and/or binding.

In more detail the PDCP may be composed two distinct elements:

a. A polynucleotide-chimeric protein complex. This links the displayed target peptide portion to the polynucleotide encoding that peptide portion through a specific polynucleotide binding portion. The nucleotide sequence encoding the chimeric protein, and the specific nucleotide sequence motif recognised by the nucleotide binding portion of the chimeric protein must be present on a segment of polynucleotide which can be incorporated into the PDCP; and

b. A protective coat. This may be supplied by a replicable carrier or helper package capable of independent existence. Alternatively, a coat protein could be encoded by the recombinant polynucleotide of the invention. The protective coat for the polynucleotide-chimeric protein complex may be composed of a biological material such as protein or lipid, but the protective coat is not required for linking the target peptide to the polynucleotide encoding that peptide. The protective coat must allow the display of the target peptide portion of the chimeric protein on its outer surface. The carrier or helper package may also provide the mechanism for releasing the intact PDCP from host cells when so required. By way of example, when a bacteriophage is the replicable carrier package, a protein coat of the bacteriophage surrounds the polynucleotide-chimeric protein complex to form the PDCP, which is then extruded from the host bacterial cell.

The invention described herein demonstrates that peptides fused to a nucleotide binding domain can be displayed externally, even through a bacteriophage carrier package protein coat, while still bound to the polynucleotide encoding the displayed peptide.

The present invention also provides a recombinant polynucleotide comprising a nucleotide sequence encoding a chimeric protein having a nucleotide binding portion operably linked to a target peptide portion, wherein said polynucleotide includes a specific nucleotide sequence motif which is bound by the nucleotide binding portion of said chimeric protein and further encoding a non-sequence-specific nucleotide binding protein.

Desirably, the recombinant polynucleotide is a recombinant expression system, able to express the chimeric protein when placed in a suitable environment, for example a compatible host cell. After its expression, the chimeric protein binds to the specific nucleotide sequence (motif) present in the polynucleotide comprising the nucleotide sequence encoding the chimeric protein.

Optionally there may be a linker sequence located between the nucleotide sequence encoding the nucleotide binding portion and the polynucleotide inserted into the restriction enzyme site of the construct.

Desirably the nucleotide binding portion is a DNA binding domain of an oestrogen or progesterone receptor, or a functional equivalent thereof. Examples of sequences encoding such nucleotide binding portions are set out in SEQ ID Nos 11 and 13.

The term "expression system" is used herein to refer to a genetic sequence which includes a protein-encoding region and is operably linked to all of the genetic signals necessary to achieve expression of that region. Optionally, the expression system may also include regulatory elements, such as a promoter or enhancer to increase transcription and/or translation of the protein encoding region or to provide control over expression. The regulatory elements may be located upstream or downstream of the protein encoding region or within the protein encoding region itself. Where two or more distinct protein encoding regions are present these may use common regulatory element(s) or have separate regulatory element(s).

Generally, the recombinant polynucleotide described above will be DNA. Where the expression system is based upon an M13 vector, usually the polynucleotide binding portion of the expressed chimeric portion will be single-stranded DNA. However, other vector systems may be used and the nucleotide binding portion may be selected to bind preferentially to double-stranded DNA or to double or single-stranded RNA, as convenient.

Additionally the present invention provides a vector containing such a recombinant expression system and host cells transformed with such a recombinant expression system (optionally in the form of a vector).

Whilst the recombinant polynucleotide described above forms an important part of the present invention, we are also concerned with the ability to screen large (e.g. of at least 10^5 members, for example 10^6 or even 10^7 members) libraries of genetic material. One of the prime considerations therefore is the provision of a recombinant genetic construct into which each member of said library can individually be incorporated to form the recombinant polynucleotide described above and to express the chimeric protein thereby encoded (the target peptide of which is encoded by the nucleotide library member incorporated into the construct).

Thus viewed in a further aspect the present invention provides a genetic construct or set of genetic constructs comprising a polynucleotide having a sequence which includes:

i) a sequence encoding a nucleotide binding portion able to recognise and bind to a specific sequence motif;
ii) the sequence motif recognised and bound by the nucleotide binding portion encoded by (i);
iii) a restriction enzyme site which permits insertion of a polynucleotide, said site being designed to operably link said polynucleotide to the sequence encoding the nucleotide binding portion so that expression of the operably linked polynucleotide sequences yields a chimeric protein; and
iv) a sequence encoding a nucleotide binding protein which binds non-specifically to naked

polynucleotide.

Optionally there may be a linker sequence located between the nucleotide sequence encoding the nucleotide binding portion and the sequence of the polynucleotide from the library inserted into the restriction enzyme site of the construct.

Desirably the nucleotide binding portion is a DNA binding domain of an oestrogen or progesterone receptor, or a functional equivalent thereof. Examples of sequences encoding such nucleotide binding portions are set out in SEQ ID Nos 11 and 13.

Suitable genetic constructs according to the invention include pDM12, pDM14 and pDM16, deposited at NCIMB on 28 August 1998 under Nos NCIMB 40970, NCIMB 40971 and NCIMB 40972 respectively.

It is envisaged that a conventionally produced genetic library may be exposed to the genetic construct(s) described above. Thus, each individual member of the genetic library will be separately incorporated into the genetic construct and the library will be present in the form of a library of recombinant polynucleotides (as described above), usually in the form of vectors, each recombinant polynucleotide including as library member.

Thus, in a further aspect, the present invention provides a library of recombinant polynucleotides (as defined above) wherein each polynucleotide includes a polynucleotide obtained from a genetic library and which encodes the target peptide portion of the chimeric protein expressed by the recombinant polynucleotide.

Optionally, the chimeric protein may further include a linker sequence located between the nucleotide binding portion and the target peptide portion. The linker sequence will reduce steric interference between the two portions of the protein. Desirably the linker sequence exhibits a degree of flexibility.

Also disclosed are methods for constructing and screening libraries of PDCP particles, displaying many different peptides, allowing the isolation and identification of particular peptides by means of affinity techniques relying on the binding activity of the peptide of interest. The resulting polynucleotide sequences can therefore be more readily identified, re-cloned and expressed.

A method of constructing a genetic library, said method comprising:

a) constructing multiple copies of a recombinant vector comprising a polynucleotide sequence which encodes a nucleotide binding portion able to recognise and bind to a specific sequence motif (and optionally also including the specific sequence motif);

b) operably linking each said vector to a polynucleotide encoding a target polypeptide, such that expression of said operably linked vector results in expression of a chimeric protein comprising said target peptide and said nucleotide binding portions; wherein said multiple copies of said operably linked vectors collectively express a library of target peptide portions;

c) transforming host cells with the vectors of step b);

d) culturing the host cells of step c) under conditions suitable for expression of said chimeric protein;

e) providing a recombinant polynucleotide comprising the nucleotide sequence motif specifically recognised by the nucleotide binding portion and exposing this polynucleotide to the chimeric protein of step d) to yield a polynucleotide-chimeric protein complex; and

f) causing production of a non-sequence-specific moiety able to bind to the non-protected portion of the polynucleotide encoding the chimeric protein to form a peptide display carrier package.

The present invention further provides a method of screening a genetic library, said method comprising:

a) exposing the polynucleotide members of said library to multiple copies of a genetic construct comprising a nucleotide sequence encoding a nucleotide binding portion able to recognise and bind to a specific sequence motif, under conditions suitable for the polynucleotides of said library each to be individually ligated into one copy of said genetic construct, to create a library of recombinant polynucleotides;

b) exposing said recombinant polynucleotides to a population of host cells, under conditions suitable for transformation of said host cells by said recombinant polynucleotides;

c) selecting for transformed host cells;

d) exposing said transformed host cells to conditions suitable for expression of said recombinant polynucleotide to yield a chimeric protein; and

e) providing a recombinant polynucleotide comprising the nucleotide sequence motif specifically recognised by the nucleotide binding portion and exposing this polynucleotide to the chimeric protein of step d) to yield a polynucleotide-chimeric protein complex;

f) protecting any exposed portions of the polynucleotide in the complex of step e) to form a peptide display carrier package; and

g) screening said peptide display carrier package to select only those packages displaying a target peptide portion having the characteristics required.

Desirably in step a) the genetic construct is pDM12, pDM14 or pDM16.

Desirably in step f) the peptide display package carrier is extruded from the transformed host cell without lysis of the host cell.

Generally the transformed host cells will be plated out or otherwise divided into single colonies following transformation and prior to expression of the chimeric protein.

The screening step g) described above may look for a particular target peptide either on the basis of function (e.g. enzymic activity) or structure (e.g. binding to a specific antibody). Once the peptide display carrier package is observed to include a target peptide with the desired characteristics, the polynucleotide portion thereof (which of course encodes the chimeric protein itself) can be amplified, cloned and otherwise manipulated using standard genetic engineering techniques.

The current invention differs from the prior art teaching of the previous disclosures US Patent No 5,403,484 and US Patent No 5,571,698, as the invention does not require outer surface transport signals, or functional portions of viral coat proteins, to enable the display of chimeric binding proteins on the outer surface of the viral particle or genetic package.

The current invention also differs from the teaching of WO-A-92/01047 and WO-A-92/20791, as no component of a secreted replicable genetic display package, or viral coat protein is required, to enable display of the target peptide on the outer surface of the viral particle.

The current invention differs from the teaching of US Patent No 5498530, as it enables the display of chimeric proteins, linked to the polynucleotide encoding the chimeric protein, extra-cellularly, not in the cytoplasm of a host cell. In the current invention the chimeric proteins are presented on the outer surface of a peptide display carrier package (PDCP) which protects the DNA encoding the chimeric protein, and does not require cell lysis to obtain access to the chimeric protein-DNA complex. Finally, the current invention does not rely upon the lacI DNA binding protein to form the chimeric protein-DNA complex.

In one embodiment of the invention, the nucleotide binding portion of the chimeric protein comprises a DNA binding domain from one or more of the nuclear steroid receptor family of proteins, or a functional equivalent of such a domain. Particular examples include (but are not limited to) a DNA binding domain of the oestrogen receptor or the progesterone receptor, or functional equivalents thereof. These domains can recognise specific DNA sequences, termed hormone response elements (HRE), which can be bound as both double and single-stranded DNA. The DNA binding domain of such nuclear steroid receptor proteins is preferred.

The oestrogen receptor is especially referred to below by way of example, for convenience since:
(a) The oestrogen receptor is a large multifunctional polypeptide of 595 amino acids which functions in the cytoplasm and nucleus of eukaryotic cells (Green et al., 1986, Science 231: 1150-1154). A minimal high affinity DNA binding domain (DBD) has been defined between amino acids 176 and 282 (Mader et al., 1993, Nucleic Acids Res. 21: 1125-1132). The functioning of this domain (i.e. DNA binding) is not inhibited by the presence of non-DNA binding domains at both the N and C terminal ends of this domain, in the full length protein.

(b) The oestrogen receptor DNA binding domain fragment (amino acids 176-282) has been expressed in *E. coli* and shown to bind to the specific double stranded DNA oestrogen receptor target HRE nucleotide sequence, as a dimer with a similar affinity (0.5nM) to the parent molecule (Murdoch et al. 1990, Biochemistry 29: 8377-8385; Mader et al., 1993, Nucleic Acids Research 21: 1125-1132). DBD dimerization on the surface of the PDCP should result in two peptides displayed per particle.

This bivalent display can aid in the isolation of low affinity peptides and peptides that are required to form a bivalent conformation in order to bind to a particular target, or activate a target receptor. The oestrogen receptor is capable of binding to its 38 base pair target HRE sequence, consensus sequence:

1) 5'-TCAGGTCAGAGTGACCTGAGCTAAAATAACACATTCAG-3'
("minus strand") SEQ ID No 1, and

2) 3'-AGTCCAGTCTCACTGGACTCGATTTTATTGTGTAAGTC-5'
("plus strand") SEQ ID No 2,

with high affinity and specificity, under the salt and pH conditions normally required for selection of binding peptides. Moreover, binding affinity is increased 60-fold for the single-stranded coding, or "plus", strand (i.e. SEQ ID No 2) of the HRE nucleotide sequence over the double stranded form of the specific target nucleotide sequence (Peale et al. 1988, Proc. Natl. Acad. Sci. USA 85: 1038-1042; Lannigan & Notides, 1989, Proc. Natl. Acad. Sci. USA 86: 863-867).

In an embodiment of the invention where the DNA binding component of the peptide display carrier package is the oestrogen receptor, the nucleotide (DNA) binding portion contains a minimum sequence of amino acids 176-282 of the oestrogen receptor protein. In addition, the consensus oestrogen receptor target HRE sequence is cloned in such a way that if single stranded DNA can be produced then the coding, or "plus", strand of the oestrogen receptor HRE nucleotide sequence is incorporated into single-stranded DNA. An example of a vector suitable for this purpose is pUC119 (see Viera et al., Methods in Enzymology, Vol 153, pages 3-11, 1987).

In a preferred embodiment of the invention a peptide display carrier package (PDCP) can be assembled when a bacterial host cell is transformed with a bacteriophage vector, which vector comprises a recombinant polynucleotide as described above. The expression vector will also comprise the specific nucleotide motif that can be bound by the nucleotide binding portion of the chimeric protein. Expression of recombinant polynucleotide results in the production of the chimeric protein which comprises the target peptide and the nucleotide binding portion. The host cell is grown under conditions suitable for chimeric protein expression and assembly of the bacteriophage particles, and the association of the chimeric protein with the specific nucleotide sequence in the expression vector.

In this embodiment, since the vector is a bacteriophage, which replicates to produce a single-stranded DNA, the nucleotide binding portion preferably has an affinity for single-stranded DNA. Incorporation of the vector single-stranded DNA-chimeric protein complex into bacteriophage particles results in the assembly of the peptide display carrier package (PDCP), and display of the target peptide on the outer surface of the PDCP.

In this embodiment both of the required elements for producing peptide display carrier packages are contained on the same vector. Incorporation of the DNA-chimeric protein complex into a peptide display carrier package (PDCP) is preferred as DNA degradation is prevented, large numbers of PDCPs are produced per host cell, and the PDCPs are easily separated from the host cell without recourse to cell lysis.

In a more preferred embodiment, the vector of the is a phagemid vector (for example pUC119) where expression of the chimeric protein is controlled by an inducible promoter. In this embodiment the PDCP can only be assembled following infection of the host cell with both phagemid vector and helper phage. The transfected host cell is then cultivated under conditions suitable for chimeric protein expression and assembly of the bacteriophage particles.

In this embodiment the elements of the PDCP are provided by two separate vectors. The phagemid derived PDCP is superior to phagemid derived display packages disclosed in WO-A-92/01047 where a proportion of packages displaying bacteriophage coat protein fusion proteins will contain the helper phage DNA, not the fusion protein DNA sequence. In the current invention, a PDCP can display the chimeric fusion protein only when the package contains the specific nucleotide motif recognised by the nucleotide binding portion. In most embodiments this sequence will be present on the same DNA segment that encodes the fusion protein. In addition, the prior art acknowledges that when mutant and wild type proteins are co-expressed in the same bacterial cell, the wild type protein is produced preferentially. Thus, when the wild type helper phage, phage display system of WO-A-92/01047 is used, both wild type gene pIII and target peptide-gene pIII chimeric proteins are produced in the same cell. The result of this is that the wild type gene pIII protein is preferentially packaged into bacteriophage particles, over the chimeric protein. In the current invention, there is no competition with wild type bacteriophage coat proteins for packaging.

Desirably the target peptide is displayed in a location exposed to the external environment of the PDCP, after the PDCP particle has been released from the host cell without recourse to cell lysis. The target peptide is then accessible for binding to its ligand. Thus, the target peptide may be located at or near the N-terminus or the C-terminus of a nucleotide binding domain, for example the DNA binding domain of the oestrogen receptor.

The present invention also provides a method for screening a DNA library expressing one or more polypeptide chains that are processed, folded and assembled in the periplasmic space to achieve biological activity. The PDCP may be assembled by the following steps:

(a) Construction of N- or C-terminal DBD chimeric protein fusions in a phagemid vector.

(i) When the target peptide is located at the N-terminus of the nucleotide binding portion, a library of DNA sequences each encoding a potential target peptide is cloned into an appropriate location of an expression vector (i.e. behind an appropriate promoter and translation sequences and a sequence encoding a signal peptide leader directing transport of the downstream fusion protein to the periplasmic space) and upstream of the sequence encoding the nucleotide binding portion. In a preferred embodiment the DNA sequence(s) of interest may be joined, by a region of DNA encoding a flexible amino acid linker, to the 5'-end of an oestrogen receptor DBD.

(ii) Alternatively, when the target peptide is located at the C-terminus of the nucleotide binding domain, a library of DNA sequences each encoding a potential target peptide is cloned into the expression vector so that the nucleotide sequence coding for the nucleotide binding portion is upstream of the cloned

DNA target peptide encoding sequences, said nucleotide binding portion being positioned behind an appropriate promoter and translation sequences and a sequence encoding a signal peptide leader directing transport of the downstream fusion protein to the periplasmic space. In a preferred embodiment, DNA sequence(s) of interest may be joined, by a region of DNA encoding a flexible amino acid linker oestrogen receptor DBD DNA sequence.

Located on the expression vector is the specific HRE nucleotide sequence recognised, and bound, by the oestrogen receptor DBD. In order to vary the number of chimeric proteins displayed on each PDCP particle, this sequence can be present as one or more copies in the vector.

(b) Incorporation into the PDCP. Non-lytic helper bacteriophage infects host cells containing the expression vector. Preferred types of bacteriophage include the filamentous phage fd, f1 and M13. In a more preferred embodiment the bacteriophage may be M13K07.

The protein(s) of interest are expressed and transported to the periplasmic space, and the properly assembled proteins are incorporated into the PDCP particle by virtue of the high affinity interaction of the DBD with the specific target nucleotide sequence present on the phagemid vector DNA which is naturally packaged into phage particles in a single-stranded form. The high affinity interaction between the DBD protein and its specific target nucleotide sequence prevents displacement by bacteriophage coat proteins resulting in the incorporation of the protein(s) of interest onto the surface of the PDCP as it is extruded from the cell.

(c) Selection of the peptide of interest. Particles which display the peptide of interest are then selected from the culture by affinity enrichment techniques. This is accomplished by means of a ligand specific for the protein of interest, such as an antigen if the protein of interest is an antibody. The ligand may be presented on a solid surface such as the surface of an ELISA plate, or in solution. Repeating the affinity selection procedure provides an enrichment of clones encoding the desired sequences, which may then be isolated for sequencing, further cloning and/or expression.

Numerous types of libraries of peptides fused to the DBD can be screened under this embodiment including:

(i) Random peptide sequences encoded by synthetic DNA of variable length.

(ii) Single-chain Fv antibody fragments. These consist of the antibody heavy and light chain variable region domains joined by a flexible linker peptide to create a single-chain antigen binding molecule.

(iii) Random fragments of naturally occurring proteins isolated from a cell population containing an activity of interest.

In another embodiment the invention concerns methods for screening a DNA library whose members require more than one chain for activity, as required by, for example, antibody Fab fragments for ligand binding. In this embodiment heavy or light chain antibody DNA is joined to a nucleotide sequence encoding a DNA binding domain of, for example, the oestrogen receptor in a phagemid vector. Typically the antibody DNA library sequences for either the heavy (VH and CH1) or light chain (VL and CL) genes are inserted in the 5′ region of the oestrogen receptor DBD DNA, behind an appropriate promoter and translation sequences and a sequence encoding a signal peptide leader directing transport of the downstream fusion protein to the periplasmic space.

Thus, a DBD fused to a DNA library member-encoded protein is produced and assembled in to the viral particle after infection with bacteriophage. The second and any subsequent chain(s) are expressed separately either:

(a) from the same phagemid vector containing the DBD and the first polypeptide fusion protein,
or

(b) from a separate region of DNA which may be present in the host cell nucleus, or on a plasmid, phagemid or bacteriophage expression vector that can co-exist, in the same host cell, with the first expression vector, so as to be transported to the periplasm where they assemble with the first chain that is fused to the DBD protein as it exits the cell. Peptide display carrier packages (PDCP) which encode the protein of interest can then be selected by means of a ligand specific for the protein.

In yet another embodiment, the invention concerns screening libraries of bi-functional peptide display carrier packages where two or more activities of interest are displayed on each PDCP. In this embodiment, a first DNA library sequence(s) is inserted next to a first DNA binding domain (DBD) DNA sequence, for example the oestrogen receptor DBD, in an appropriate vector, behind an appropriate promoter and translation sequences and a sequence encoding a signal peptide leader directing transport of this first chimeric protein to the periplasmic space. A second chimeric protein is also produced from the same, or separate, vector by inserting a second DNA library sequence(s) next to a second DBD DNA sequence which is different from the first DBD DNA sequence, for example the progesterone receptor DBD, behind an appropriate promoter and translation sequences and a sequence encoding a signal peptide leader. The first, or only, vector contains the specific HRE nucleotide sequences for both oestrogen and progesterone receptors. Expression of the two chimeric proteins, results in a PDCP with two different chimeric proteins displayed. As an example, one chimeric protein could possess a binding activity for a particular ligand of interest, while the second chimeric protein could possess an enzymatic activity. Binding by the PDCP to the ligand of the first chimeric protein could then be detected by subsequent incubation with an appropriate substrate for the second chimeric protein. In an alternative embodiment a bi-functional PDCP may be created using a single DBD, by cloning one peptide at the 5'-end of the DBD, and a second peptide at the 3'-end of the DBD. Expression of this single bi-functional chimeric protein results in a PDCP with two different activities.

We have investigated the possibility of screening libraries of peptides, fused to a DNA binding domain and displayed on the surface of a display package, for particular peptides with a biological activity of interest and recovering the DNA encoding that activity. Surprisingly, by manipulating the oestrogen receptor

DNA binding domain in conjunction with M13 bacteriophage we have been able to construct novel particles which display large biologically functional molecules, that allows enrichment of particles with the desired specificity.

The invention described herein provides a significant breakthrough in DNA library screening technology.

The invention will now be further described by reference to the non-limiting examples and figures below.

Description of Figures

Figure 1 shows the pDM12 N-terminal fusion oestrogen receptor DNA binding domain expression vector nucleotide sequence (SEQ ID No 3), between the HindIII and EcoRI restriction sites, comprising a pelB leader secretion sequence (in italics), multiple cloning site containing SfiI and NotI sites, flexible (glycine)$_4$-serine linker sequence (boxed), a fragment of the oestrogen receptor gene comprising amino acids 176-282 (SEQ ID No 4) of the full length molecule, and the 38 base pair consensus oestrogen receptor DNA binding domain HRE sequence.

Figure 2 shows the OD_{450nm} ELISA data for negative control M13K07 phage, and single-clone PDCP display culture supernatants (#1-4, see Example 3) isolated by selection of the lymphocyte cDNA-pDM12 library against anti-human immunoglobulin kappa antibody.

Figure 3 shows partial DNA (SEQ ID No 5) and amino acid (SEQ ID No 6) sequence for the human immunoglobulin kappa constant region (Kabat, E. A. et al., Sequences of Proteins of Immunological Interest. 4th edition. U.S. Department of Health and Human Services. 1987), and ELISA positive clones #2 (SEQ ID Nos 7 and 8) and #3 (SEQ ID Nos 9 and 10) from Figure 2 which confirms the presence of human kappa constant region DNA in-frame with the pelB leader sequence (pelB leader sequence is underlined, the leader sequence cleavage site is indicated by an arrow). The differences in the 5'-end sequence demonstrates that these two clones were selected independently from the library stock. The PCR primer sequence is indicated in bold, clone #2 was originally amplified with CDNAPCRBAK1 and clone #3 was amplified with CDNAPCRBAK2.

Figure 4 shows the pDM14 N-terminal fusion oestrogen receptor DNA binding domain expression vector nucleotide sequence (SEQ ID No 11), between the HindIII and EcoRI restriction sites, comprising a pelB leader secretion sequence (in italics), multiple cloning site containing SfiI and NotI sites, flexible (glycine)$_4$-serine linker sequence (boxed), a fragment of the oestrogen receptor gene comprising amino acids 176-282 (SEQ ID No 12) of the full length molecule, and the two 38 base pair oestrogen receptor DNA binding domain HRE sequences (HRE 1 and HRE 2).

Figure 5 shows the pDM16 C-terminal fusion oestrogen receptor DNA binding domain expression vector nucleotide sequence (SEQ ID No 13), between the HindIII and EcoRI restriction sites, comprising a pelB leader secretion sequence (in italics), a fragment of the oestrogen receptor gene comprising amino acids 176-282 (SEQ ID No 14) of the full length molecule, flexible (glycine)$_4$-serine linker sequence (boxed), multiple cloning site containing SfiI and NotI sites and the 38 base pair oestrogen receptor DNA binding domain HRE sequence.

Figure 6 shows the OD_{450nm} ELISA data for N-cadherin-pDM16 C-terminal display PDCP binding to anti-pan-cadherin monoclonal antibody in serial dilution ELISA as ampicillin resitance units (a.r.u.). Background binding of negative control M13K07 helper phage is also shown.

Figure 7 shows the OD_{450nm} ELISA data for *in vivo* biotinylated PCC-pDM16 C-terminal display PDCP binding to streptavidin in serial dilution ELISA as ampicillin resitance units (a.r.u.). Background binding of negative control M13K07 helper phage is also shown.

Figure 8 shows the OD_{450nm} ELISA data for a human scFv PDCP isolated from a human scFv PDCP display library selected against substance P. The PDCP was tested against streptavidin (1), streptavidin-biotinylated substance P (2), and streptavidin-biotinylated CGRP (3), in the presence (B) or absence (A) of free substance P.

Figure 9 shows the DNA (SEQ ID Nos 15 and 17) and amino acid (SEQ ID No 16 and 18) sequence of the substance P binding scFv isolated from a human scFv PDCP display library selected against substance P. Heavy chain (SEQ ID Nos 15 and 16) and light chain (SEQ ID Nos 17 and 18) variable region sequence is shown with the CDRs underlined and highlighted in bold.

Materials and Methods

The following procedures used by the present applicant are described in Sambrook, J., et al., 1989 supra.: restriction enzyme digestion, ligation, preparation of electrocompetent cells, electroporation, analysis of restriction enzyme digestion products on agarose gels, DNA purification using phenol/chloroform, preparation of 2xTY medium and plates, preparation of ampicillin, kanamycin, IPTG (Isopropyl β-D-Thiogalactopyranoside) stock solutions, and preparation of phosphate buffered saline.

Restriction enzymes, T4 DNA ligase and cDNA synthesis reagents (Superscript plasmid cDNA synthesis kit) were purchased from Life Technologies Ltd (Paisley, Scotland, U.K.). Oligonucleotides were obtained from Cruachem Ltd (Glasgow, Scotland, U.K.), or Genosys Biotechnologies Ltd (Cambridge, U.K.). Taq DNA polymerase, Wizard SV plasmid DNA isolation kits, streptavidin coated magnetic beads and mRNA isolation reagents (PolyATract 1000) were obtained from Promega Ltd (Southampton, Hampshire, U.K.). Taqplus DNA polymerase was obtained from Stratagene Ltd (Cambridge, U.K.). PBS, BSA, streptavidin, substance P and anti-pan cadherin antibody were obtained from SIGMA Ltd (Poole, Dorset, U.K.). Anti-M13-HRP conjugated antibody, Kanamycin resistant M13K07 helper bacteriophage and RNAguard were obtained from Pharmacia Ltd (St. Albans, Herts, U.K.) and anti-human Igκ antibody from Harlan-Seralab (Loughborough, Leicestershire, U.K.) Biotinylated substance P and biotinylated calcitonin gene related peptide (CGRP) were obtained from Peninsula Laboratories (St. Helens, Merseyside, U.K.).

Specific embodiments of the invention are given below in Examples 1 to 9.

Example 1. Construction of a N-terminal PDCP display phagemid vector pDM12.

The pDM12 vector was constructed by inserting an oestrogen receptor DNA binding domain, modified by appropriate PCR primers, into a phagemid vector pDM6. The pDM6 vector is based on the pUC119 derived phage display vector pHEN1 (Hoogenboom et al., 1991, Nucleic Acids Res. 19: 4133-4137). It contains $(Gly)_4$Ser linker, Factor Xa cleavage site, a full length gene III, and streptavidin tag peptide sequence (Schmidt, T.G. and Skerra, A., 1993, Protein Engineering 6: 109-122), all of which can be removed by NotI-EcoRI digestion and agarose gel electrophoresis, leaving a pelB leader sequence, SfiI, NcoI and PstI restriction sites upstream of the digested NotI site. The cloned DNA binding domain is under the control of the lac promoter found in pUC119.

Preparation of pDM6

The pDM12 vector was constructed by inserting an oestrogen receptor DNA binding domain, modified by appropriate PCR primers, into a phagemid vector pDM6. The pDM6 vector is based on the gene pIII phage display vector pHEN1 (Hoogenboom et al., 1991, Nucleic Acids Res. 19: 4133-4137), itself derived from pUC119 (Viera, J. and Messing, J., 1987, Methods in Enzymol. 153: 3-11). It was constructed by amplifying the pIII gene in pHEN1 with two oligonucleotides:

PDM6BAK: 5 -TTT TCT GCA GTA ATA GGC GGC CGC AGG GGG AGG AGG GTC CAT CGA AGG TCG CGA AGC AGA GAC TGT TGA AAG T-3 (SEQ ID No 19) and

PDM6FOR: 5 - TTT TGA ATT CTT ATT AAC CAC CGA ACT GCG

GGT GAC GCC AAG CGC TTG CGG CCG TTA AGA CTC CTT ATT ACG CAG-3 (SEQ ID No 20).

and cloning the PstI-EcoRI digested PCR product back into similarly digested pHEN1, thereby removing the c-myc tag sequence and supE TAG codon from pHEN1. The pDM6 vector contains a $(Gly)_4$Ser linker, Factor Xa cleavage site, a full length gene III, and streptavidin tag peptide sequence (Schmidt, T.G. and Skerra, A., 1993, Protein Engineering 6: 109-122), all of which can be removed by NotI-EcoRI digestion and agarose gel electrophoresis, leaving a pelB leader sequence, SfiI, NcoI and PstI restriction sites upstream of the digested NotI site. The cloned DNA binding domain is under the control of the lac promoter found in pUC119.

The oestrogen receptor DNA binding domain was isolated from cDNA prepared from human bone marrow (Clontech, Palo Alto, California, U.S.A.). cDNA can be prepared by many procedures well known to those skilled in the art. As an example, the following method using a Superscript plasmid cDNA synthesis kit can be used:

(a) First strand synthesis.

5μg of bone marrow mRNA, in 5μl DEPC-treated water was thawed on ice and 2μl (50pmol) of cDNA synthesis primer (5'-AAAAGCGGCCGCACTGGCCTGAGAGA$(N)_6$-3') (SEQ ID No 21) was added to the mRNA and the mixture heated to 70°C for 10 minutes, then snap-chilled on ice and spun briefly to collect the contents to the bottom of the tube. The following were then added to the tube:

1000u/ml RNAguard	1μl
5x first strand buffer	4μl
0.1M DTT	2μl
10mM dNTPs	1μl

200u/μl SuperScript II reverse transcriptase 5μl

The mixture was mixed by pipetting gently and incubated at 37°C for 1 hour, then placed on ice.

(b) Second strand synthesis.

The following reagents were added to the first strand reaction:

DEPC-treated water	93μl
5x second strand buffer	30μl
10mM dNTPs	3μl
10u/μl *E. coli* DNA ligase	1μl
10u/μl *E. coli* DNA polymerase	4μl
2u/μl *E. coli* RNase H	1μl

The reaction was vortex mixed and incubated at 16°C for 2 hours. 2μl (10u) of T4 DNA polymerase was added and incubation continued at 16°C for 5 minutes. The reaction was placed on ice and 10μl 0.5M EDTA added, then phenol-chloroform extracted, precipitated and vacuum dried.

(c) Sal I adaptor ligation.

The cDNA pellet was resuspended in 25μl DEPC-treated water, and ligation set up as follows.

cDNA	25μl
5x T4 DNA ligase buffer	10μl
1μg/μl *Sal* I adapters*	10μl
1u/μlT4 DNA ligase	5μl

*Sal I adapters: TCGACCCACGCGTCCG-3' (SEQ ID No 22)
GGGTGCCGAGGC-5' (SEQ ID No 23)

The ligation was mixed gently and incubated for 16 hours at 16°C, then phenol-chloroform extracted, precipitated and vacuum dried. The cDNA/adaptor pellet was resuspended in 41μl of DEPC-treated water and digested with 60 units of NotI at 37°C for 2 hours, then phenol-chloroform extracted, precipitated and vacuum dried. The cDNA pellet was re-dissolved in 100μl TEN buffer (10mM Tris pH 7.5, 0.1mM EDTA, 25mM NaCl) and size fractionated using a Sephacryl S-500 HR column to remove unligated adapters and small cDNA fragments (<400bp) according to the manufacturers instructions. Fractions were checked by agarose gel electrophoresis and fractions containing cDNA less than 400 base pairs discarded, while the remaining fractions were pooled.

(d) PCR amplification of oestrogen receptor DNA binding domain.

The oestrogen receptor was PCR amplified from 5μl (150-250ng) of bone marrow cDNA using 25pmol of each of the primers pDM12FOR (SEQ ID No 24) (5'-AAAAGAATTCTGAATGTGTTATTTAGCTCAGGTCACTCTGACCTGATTATCAAGACCCCACTTCACCCCCT) and pDM12BAK (SEQ ID No 25) (5'-AAAAGCGGCCGCAGGGGGAGGAGGGTCCATGGAATCTGCCAAGGAG-3') in two 50μl reactions containing 0.1mM dNTPs, 2.5 units Taq DNA polymerase, and 1x PCR reaction buffer (10mM Tris-HCl pH 9.0, 5mM KCl, 0.01% Triton X®-100, 1.5mM $MgCl_2$) (Promega Ltd, Southampton, U.K.). The pDM12FOR primer anneals to the 3'-end of the DNA binding domain of the oestrogen receptor and incorporates two stop codons, the 38 base pair consensus oestrogen receptor HRE sequence, and an EcoRI restriction site. The pDM12BAK primer anneals to the 5'-end of the DNA binding domain of the oestrogen receptor and incorporates the $(Gly)_4Ser$ linker and the NotI restriction site.

Reactions were overlaid with mineral oil and PCR carried out on a Techne PHC-3 thermal cycler for 30 cycles of 94°C, 1 minute; 65°C, 1 minute; 72°C, 1 minute. Reaction products were electrophoresed on an agarose gel, excised and products purified from the gel using a Geneclean II kit according to the manufacturers instructions (Bio101, La Jolla, California, U.S.A.).

(e) Restriction digestion and ligation.

The PCR reaction appended NotI and EcoRI restriction sites, the $(Gly)_4Ser$ linker, stop codons and the 38 base pair oestrogen receptor target HRE nucleotide sequence to the oestrogen receptor DNA binding domain sequence (see Figure 1). The DNA PCR fragment and the target pDM6 vector (approximately 500ng) were NotI and EcoRI digested for 1 hour at 37°C, and DNA purified by agarose gel electrophoresis and extraction with Geneclean II kit (Bio101, La Jolla, California, U.S.A.). The oestrogen receptor DNA binding domain cassette was ligated into the NotI-EcoRI digested pDM6 vector overnight at 16°C, phenol/chloroform extracted and precipitated then electroporated into TG1 *E. coli* (genotype: K12, (Δlac-pro), supE, thi, hsD5/F'traD36, $proA^+B^+$, $LacI^q$, LacZΔ15) and plated onto 2xTY agar plates supplemented with 1% glucose and 100μg/ml ampicillin. Colonies were allowed to grow overnight at 37°C. Individual colonies were picked into 5ml 2xTY supplemented with 1% glucose and 100μg/ml ampicillin and grown overnight at 37°C. Double stranded phagemid DNA was isolated with a Wizard SV plasmid DNA isolation kit and the sequence confirmed with a Prism dyedeoxy cycle sequencing kit (Perkin-Elmer, Warrington, Lancashire, U.K.) using M13FOR (SEQ ID No 26) (5'-GTAAAACGACGGCCAGT) and M13REV (SEQ ID No 27) (5'-GGATAACAATTTCACACAGG) oligonucleotides. The pDM12 PDCP display vector DNA sequence between the HindIII and EcoRI restriction sites is shown in Figure 1.

Example 2. Insertion of a random-primed human

lymphocyte cDNA into pDM12 and preparation of a master PDCP stock.

Libraries of peptides can be constructed by many methods known to those skilled in the art. The example given describes a method for constructing a peptide library from randomly primed cDNA, prepared from mRNA isolated from a partially purified cell population.

mRNA was isolated from approximately 10^9 human peripheral blood lymphocytes using a polyATract 1000 mRNA isolation kit (Promega, Southampton, UK). The cell pellet was resuspended in 4ml extraction buffer (4M guanidine thiocyanate, 25mM sodium citrate pH 7.1, 2% β-mercaptoethanol). 8ml of pre-heated (70°C) dilution buffer (6xSSC, 10mM Tris pH 7.4, 1mM EDTA, 0.25% SDS, 1% β-mercaptoethanol) was added to the homogenate and mixed thoroughly by inversion. 10μl of biotinylated oligo-dT (50 pmol/μl) was added, mixed and the mixture incubated at 70°C for 5 minutes. The lymphocyte cell lysate was transferred to 6x 2ml sterile tubes and spun at 13,000 rpm in a microcentrifuge for ten minutes at ambient temperature to produce a cleared lysate. During this centrifugation, streptavidin coated magnetic beads were resuspended and 6ml transferred to a sterile 50ml Falcon tube, then placed in the magnetic stand in a horizontal position until all the beads were captured. The supernatant was carefully poured off and beads resuspended in 6ml 0.5xSSC, then the capture repeated. This wash was repeated 3 times, and beads resuspended in a final volume of 6ml 0.5xSSC. The cleared lysate was added to the washed beads, mixed by inversion and incubated at ambient temperature for 2 minutes, then beads captured in the magnetic stand in a horizontal position. The beads were resuspended gently in 2ml 0.5xSSC and transferred to a sterile 2ml screwtop tube, then captured again in the vertical position, and the wash solution discarded. This wash was repeated twice more. 1ml of DEPC-treated water was added to the beads and mixed gently. The beads were again captured and the eluted mRNA transferred to a sterile tube. 50μl was electrophoresed to check the quality and quantity of mRNA, while the remainder was precipitated with 0.1 volumes 3M sodium acetate and three volumes absolute ethanol at -80°C overnight in 4 aliquots in sterile 1.5ml screwtop tubes.

Double stranded cDNA was synthesised as described in Example 1 using 5μg of lymphocyte mRNA as template. The cDNA was PCR amplified using oligonucleotides CDNAPCRFOR (SEQ ID No 28) (5'-AAAGCGGCCGCACTGGCCTGAGAGA), which anneals to the cDNA synthesis oligonucleotide described in Example 1 which is present at the 3'-end of all synthesised cDNA molecules incorporates a NotI restriction site, and an equimolar mixture of CDNAPCRBAK1, CDNAPCRBAK2 and CDNAPCRBAK3.
CDNAPCRBAK1: (SEQ ID No 29) 5'-AAAAGGCCCAGCCGGCCATGGCCCAGCCCACCACGCGTCCG,
CDNAPCRBAK2: (SEQ ID No 30) 5'-AAAAGGCCCAGCCGGCCATGGCCCAGTCCCACCACGCGTCCG,
CDNAPCRBAK3: (SEQ ID No 31) 5'-AAAAGGCCCAGCCGGCCATGGCCCAGTACCCACCACGCGTCCG),
all three of which anneal to the SalI adaptor sequence found at the 5'-end of the cDNA and incorporate a SfiI restriction site at the cDNA 5'-end. Ten PCR reactions were carried out using 2μl of cDNA (50ng) per reaction as described in Example 1 using 25 cycles of 94°C, 1 minute; 60°C, 1 minute; 72°C, 2 minutes. The reactions were pooled and a 20μl aliquot checked by agarose gel electrophoresis, the remainder was phenol/chloroform extracted and ethanol precipitated and resuspended in

100μl sterile water. 5μg of pDM12 vector DNA and lymphocyte cDNA PCR product were SfiI-NotI digested phenol/chloroform extracted and small DNA fragments removed by size selection on Chromaspin 1000 spin columns (Clontech, Palo Alto, California, U.S.A.) by centrifugation at 700g for 2 minutes at room temperature. Digested pDM12 and lymphocyte cDNA were ethanol precipitated and ligated together for 16 hours at 16°C. The ligated DNA was precipitated and electroporated in to TG1 *E. coli*. Cells were grown in 1ml SOC medium per cuvette used for 1 hour at 37°C, and plated onto 2xTY agar plates supplemented with 1% glucose and 100μg/ml ampicillin. 10^{-4}, 10^{-5} and 10^{-6} dilutions of the electroporated bacteria were also plated to assess library size. Colonies were allowed to grow overnight at 30°C. 2×10^{8} ampicillin resistant colonies were recovered on the agar plates.

The bacteria were then scraped off the plates into 40ml 2xTY broth supplemented with 20% glycerol, 1% glucose and 100μg/ml ampicillin. 5ml was added to a 20ml 2xTY culture broth supplemented with 1% glucose and 100μg/ml ampicillin and infected with 10^{11} kanamycin resistance units (kru) M13K07 helper phage at 37°C for 30 minutes without shaking, then for 30 minutes with shaking at 200rpm. Infected bacteria were transferred to 200ml 2xTY broth supplemented with 25μg/ml kanamycin, 100μg/ml ampicillin, and 20μM IPTG, then incubated overnight at 37°C, shaking at 200rpm. Bacteria were pelleted at 4000rpm for 20 minutes in 50ml Falcon tubes, and 40ml 2.5M NaCl/20% PEG 6000 was added to 200ml of particle supernatant, mixed vigorously and incubated on ice for 1 hour to precipitate PDCP particles. Particles were pelleted at 11000rpm for 30 minutes in 250ml Oakridge tubes at 4°C in a Sorvall RC5B centrifuge, then resuspended in 2ml PBS buffer after removing all traces of PEG/NaCl with a pipette, then bacterial debris removed by a 5 minute 13500rpm spin in a microcentrifuge. The supernatent was filtered through a 0.45μm polysulfone syringe filter and stored at -20°C.

Example 3. Isolation of human immunoglobulin kappa light chains by repeated rounds of selection against anti-human kappa antibody.

For the first round of library selection a 70x11mm NUNC Maxisorp Immunotube (Life Technologies, Paisley, Scotland U.K.) was coated with 2.5ml of 10μg/ml of anti-human kappa antibody (Seralab, Crawley Down, Sussex, U.K.) in PBS for 2 hours at 37°C. The tube was rinsed three times with PBS (fill & empty) and blocked with 3ml PBS/2% BSA for 2 hours at 37°C and washed as before. 4×10^{12} a.r.u. of pDM12-lymphocyte cDNA PDCP stock was added in 2ml 2% BSA/PBS/0.05% Tween 20, and incubated for 30 minutes on a blood mixer, then for 90 minutes standing at ambient temperature. The tube was washed ten times with PBS/0.1% Tween 20, then a further ten times with PBS only. Bound particles were eluted in 1ml of freshly prepared 0.1M triethylamine for 10 minutes at ambient temperature on a blood mixer. Eluted particles were transferred to 0.5ml 1M Tris pH 7.4, vortex mixed briefly and transferred to ice.

Neutralised particles were added to 10ml log phase TG1 E coli bacteria (optical density: OD_{600nm} 0.3-0.5) and incubated at 37°C without shaking for 30 minutes, then with shaking at 200rpm for 30 minutes. 10^{-3}, 10^{-4} & 10^{-5} dilutions of the infected culture were prepared to estimate the number of particles recovered, and the remainder was spun at 4000 rpm for 10 minutes, and the pellet resuspended in 300μl 2xTY medium by vortex mixing. Bacteria were plated onto 2xTY agar plates supplemented with 1% glucose and 100μg/ml ampicillin. Colonies were allowed to grow overnight at 30°C.

A PDCP stock was prepared from the bacteria recovered from the first round of selection, as described in Example 2 from a 100ml overnight culture. 250μl of the round 1 amplified PDCP stock was then selected against anti-human kappa antibody as described above with the tube was washed twelve times with PBS/0.1% Tween 20, then a further twelve times with PBS only.

To identify selected clones, eighty-eight individual clones recovered from the second round of selection were then tested by ELISA for binding to anti-human kappa antibody. Individual colonies were picked into 100μl 2xTY supplemented with 100μg/ml ampicillin and 1% glucose in 96-well plates (Costar) and incubated at 37°C and shaken at 200rpm for 4 hours. 25μl of each culture was transferred to a fresh 96-well plate, containing 25μl/well of the same medium plus 10^7 k.r.u. M13K07 kanamycin resistant helper phage and incubated at 37°C for 30 minutes without shaking, then incubated at 37°C and shaken at 200rpm for a further 30 minutes. 160μl of 2xTY supplemented with 100μg/ml ampicillin, 25μg/ml kanamycin, and 20μM IPTG was added to each well and particle amplification continued for 16 hours at 37°C while shaking at 200rpm. Bacterial cultures were spun in microtitre plate carriers at 2000g for 10 minutes at 4°C in a benchtop centrifuge to pellet bacteria and culture supernatant used for ELISA.

A Dynatech Immulon 4 ELISA plate was coated with 200ng/well anti-human kappa antibody in 100μl /well PBS for one hour at 37°C. The plate was washed 2x200μl/well PBS and blocked for 1 hour at 37°C with 200μl/well 2% BSA/PBS and then washed 2x200μl/well PBS. 50μl PDCP culture supernatant was added to each well containing 50μl/well 4% BSA/PBS/0.1%Tween 20, and allowed to bind for 1 hour at ambient temperature. The plate was washed three times with 200μl/well PBS/0.1% Tween 20, then three times with 200μl/well PBS. Bound PDCPs were detected with 100μl/well, 1:5000 diluted anti-M13-HRP conjugate (Pharmacia) in 2% BSA/PBS/0.05% Tween 20 for 1 hour at ambient temperature and the plate washed six times as above. The plate was developed for 5 minutes at ambient temperature with 100μl/well freshly prepared TMB (3,3',5,5'-Tetramethylbenzidine) substrate buffer (0.005% H_2O_2, 0.1mg/ml TMB in 24mM citric acid/52mM sodium phosphate buffer pH 5.2). The reaction was stopped with 100μl/well 12.5% H_2SO_4 and read at 450nm. (ELISA data for binding clones is shown in Figure 2).

These clones were then sequenced with M13REV primer (SEQ ID No 27) as in Example 1. The sequence of two of the clones isolated is shown in Figure 3 (see SEQ ID Nos 7 to 10).

Example 4. Construction of the pDM14 N-terminal display vector

It would be useful to design vectors that contain a second DBD binding sequence, such as a second oestrogen receptor HRE sequence, thus allowing the display of increased numbers of peptides per PDCP. Peale et al. (1988, Proc. Natl. Acad. Sci. USA 85: 1038-1042) describe a number of oestrogen receptor HRE sequences. These sequences were used to define an HRE sequence, which differs from that cloned in pDM12, which we used to create a second N-terminal display vector (pDM14). The oligonucleotide: 5'-AAAAGAATTCGAGGTTACATTAACTTTGTT CCGGTCAGACTGACCCAAGTCGACCTGAATGTGTTATTTTAG-3' (SEQ ID No 32) was synthesised and used to mutagenise pDM12 by

PCR with pDM12BAK oligonucleotide as described in Example 1 using 100ng pDM12 vector DNA as template. The resulting DNA fragment, which contained the oestrogen receptor DBD and two HRE sequences separated by a SalI restriction enzyme site, was NotI-EcoRI restriction enzyme digested and cloned into NotI-EcoRI digested pDM12 vector DNA as described in Example 1 to create pDM14. The sequence of pDM14 between the HindIII and EcoRI restriction enzyme sites was checked by DNA sequencing. The final vector sequence between these two sites is shown in Figure 4 (see SEQ ID Nos 11 and 12).

Example 5. Construction of the pDM16 C-terminal display vector

In order to demonstrate the display of peptides fused to the C-terminus of a DBD on a PDCP a suitable vector, pDM16, was created.

In pDM16 the pelB leader DNA sequence is fused directly to the oestrogen receptor DBD sequence removing the multiple cloning sites and the Gly_4Ser linker DNA sequence found in pDM12 and pDM14, which are appended to the C-terminal end of the DBD sequence upstream of the HRE DNA sequence.

To create this vector two separate PCR reactions were carried out on a Techne Progene thermal cycler for 30 cycles of 94°C, 1 minute; 60°C, 1 minute; 72°C, 1 minute. Reaction products were electrophoresed on an agarose gel, excised and products purified from the gel using a Mermaid or Geneclean II kit, respectively, according to the manufacturers instructions (Bio101, La Jolla, California, U.S.A.).

In the first, the 5'-untranslated region and pelB leader DNA sequence was amplified from 100ng of pDM12 vector DNA using 50pmol of each of the oligonucleotides pelBFOR (SEQ ID No 33) (5'-CCTTGGCAGATTCCATCT CGGCCATTGCCGGC-3') and M13REV (SEQ ID NO 27) (see above) in a 100μl reaction containing 0.1mM dNTPs, 2.5 units Taqplus DNA polymerase, and 1x High Salt PCR reaction buffer (20mM Tris-HCl pH 9.2, 60mM KCl, 2mM $MgCl_2$) (Stratagene Ltd, Cambridge, U.K.).

In the second, the 3'-end of the pelB leader sequence and the oestrogen receptor DBD was amplified from 100ng of pDM12 vector DNA using 50pmol of each of the oligonucleotides pelBBAK (SEQ ID No 34) (5'-CCGGCAA TGGCCGAGATGGAATCTGCCAAGG-3') and pDM16FOR (SEQ ID No 35) (5'-TTTTGTCGACTCAATCAGTTATGCGGCCGCCAGCTGCAGG AGGGCCGGCTGGGCCGACCCTCCTCCCCCAGACCCCACTTCACCCC-3') in a 100μl reaction containing 0.1mM dNTPs, 2.5 units Taqplus DNA polymerase, and 1x High Salt PCR reaction buffer (Stratagene Ltd, Cambridge, U.K.). Following gel purification both products were mixed together and a final round of PCR amplification carried out to link the two products together as described above, in a 100μl reaction containing 0.1mM dNTPs, 2.5 units Taq DNA polymerase, and 1x PCR reaction buffer (10mM Tris-HCl pH 9.0, 5mM KCl, 0.01% Triton X*-100, 1.5mM $MgCl_2$) (Promega Ltd, Southampton, U.K.).

The resulting DNA fragment, was HindIII-SalI restriction enzyme digested and cloned into HindIII-SalI digested pDM14 vector DNA as described in Example 1 to create pDM16. The sequence of pDM16 between the HindIII and EcoRI restriction enzyme sites was checked by DNA sequencing. The final vector sequence between these two sites is shown in Figure 5 (see SEQ ID Nos 13 and 14).

Example 6. Display of the C-terminal fragment of human N-cadherin on the surface of a PDCP

cDNA libraries of peptides can be constructed by many methods known to those skilled in the art. One commonly used method for constructing a peptide library uses oligo dT primed cDNA, prepared from polyA+ mRNA. In this method the first-strand synthesis is carried out using an oligonucleotide which anneals to the 3'-end polyA tail of the mRNA composed of T_n (where n is normally between 10 and 20 bases) and a restriction enzyme site such as NotI to facilitate cloning of cDNA. The cDNA cloned by this method is normally composed of the polyA tail, the 3'- end untranslated region and the C-terminal coding region of the protein. As an example of the C-terminal display of peptides on a PDCP, a human cDNA isolated from a library constructed by the above method was chosen.

The protein N-cadherin is a cell surface molecule involved in cell-cell adhesion. The C-terminal cytoplasmic domain of the human protein (Genbank database accession number: M34064) is recognised by a commercially available monoclonal antibody which was raised against the C-terminal 23 amino acids of chicken N-cadherin (SIGMA catalogue number: C-1821). The 1.4kb human cDNA fragment encoding the C-terminal 99 amino acids, 3'- untranslated region and polyA tail (NotI site present at the 3'-end of the polyA tail) was amplified from approximately 20ng pDM7-NCAD#C with 25pmol of each oligonucleotide M13FOR (SEQ ID No 26) and CDNPCRBAK1 (SEQ ID No 29) (see above) in a 50μl reaction containing 0.1mM dNTPs, 2.5 units Taqplus DNA polymerase, and 1x High Salt PCR reaction buffer (20mM Tris-HCl pH 9.2, 60mM KCl, 2mM $MgCl_2$) (Stratagene Ltd, Cambridge, U.K.) on a Techne Progene thermal cycler for

30 cycles of 94°C, 1 minute; 60°C, 1 minute; 72°C, 1 minute. Following gel purification and digestion with SfiI and NotI restiction enzymes, the PCR product was cloned into pDM16 using an analogous protocol as described in Example 1.

Clones containing inserts were identified by ELISA of 96 individual PDCP cultures prepared as described in Example 3. A Dynatech Immulon 4 ELISA plate was coated with 1:250 diluted anti-pan cadherin monoclonal antibody in 100μl /well PBS overnight at 4ºC. The plate was washed 3x200μl/well PBS and blocked for 1 hour at 37°C with 200μl/well 2% Marvel non-fat milk powder/PBS and then washed 2x200μl/well PBS. 50μl PDCP culture supernatant was added to each well containing 50μl/well 4% Marvel/PBS, and allowed to bind for 1 hour at ambient temperature. The plate was washed three times with 200μl/well PBS/0.1% Tween 20, then three times with 200μl/well PBS. Bound PDCPs were detected with 100μl/well, 1:5000 diluted anti-M13-HRP conjugate (Pharmacia) in 2% Marvel/PBS for 1 hour at ambient temperature and the plate washed six times as above. The plate was developed for 15 minutes at ambient temperature with 100μl/well freshly prepared TMB (3,3',5,5'-Tetramethylbenzidine) substrate buffer (0.005% H_2O_2, 0.1mg/ml TMB in 24mM citric acid/52mM sodium phosphate buffer pH 5.2). The reaction was stopped with 100μl/well 12.5% H_2SO_4 and read at 450nm. The nucleotide sequence of an ELISA positive clone insert and DBD junction was checked by DNA sequencing using oligonucleotides M13FOR (SEQ ID No 26) (see Example 1) and ORSEQBAK (SEQ ID No 36) (5'-TGTTGAAACACAAGCGCCAG-3').

A fifty-fold concentrated stock of C-terminal N-cadherin PDCP particles was prepared by growing the uninfected TG1 clone in 1ml 2xTY culture broth supplemented with 1% glucose and 100μg/ml ampicillin for five hours at 37°C, shaking at 200rpm and infecting with 10^8 kanamycin resistance units (kru) M13K07 helper phage at 37°C for 30 minutes without shaking, then for 30 minutes with shaking at 200rpm. Infected bacteria were transferred to 20ml 2xTY broth supplemented with 25μg/ml kanamycin, 100μg/ml ampicillin, and 20μM IPTG, then incubated overnight at 30°C, shaking at 200rpm. Bacteria were pelleted at 4000rpm for 20 minutes in 50ml Falcon tubes, and 4ml 2.5M NaCl/20% PEG 6000 was added to 20ml of PDCP supernatant, mixed vigorously and incubated on ice for 1 hour to precipitate particles.

The particles were pelleted at 11000rpm for 30 minutes in 50ml Oakridge tubes at 4°C in a Sorvall RC5B centrifuge, then resuspended in PBS buffer after removing all traces of PEG/NaCl with a pipette, then bacterial debris removed by a 5 minute 13500rpm spin in a microcentrifuge. The supernatant was filtered through a 0.45μm polysulfone syringe filter. The concentrated stock was two-fold serially diluted and used in ELISA against plates coated with anti-pan-cadherin antibody as described above (see Figure 6).

This example demonstrates the principle of C-terminal display using PDCPs, that C-terminal DBD-peptide fusion PDCPs can be made which can be detected in ELISA, and the possibility that oligo dT primed cDNA libraries may be displayed using this method.

Example 7. Display of *in vivo* biotinylated C-terminal domain of human propionyl CoA carboxylase on the surface of a PDCP

Example 6 shows that the C-terminal domain of human N-cadherin can be expressed on the surface of a PDCP as a C-terminal fusion with the DBD. Here it is shown that the C-terminal domain of another human protein propionyl CoA carboxylase alpha chain (Genbank accession number: X14608) can similarly be displayed, suggesting that this methodology may be general.

The alpha sub-unit of propionyl CoA carboxylase alpha chain (PCC) contains 703 amino acids and is normally biotinylated at position 669. It is demonstrated that the PCC peptide displayed on the PDCP is biotinylated, as has been shown to occur when the protein is expressed in bacterial cells (Leon-Del-Rio & Gravel; 1994, J. Biol. Chem. 37, 22964-22968).

The 0.8kb human cDNA fragment of PCC alpha encoding the C-terminal 95 amino acids, 3'- untranslated region and polyA tail (NotI site present at the 3'-end of the polyA tail) was amplified and cloned into pDM16 from approximately 20ng pDM7-PCC#C with 25pmol of each oligonucleotide M13FOR (SEQ ID No 26) and CDNPCRBAK1 (SEQ ID No 29) as described in Example 6.

Clones containing inserts were identified by ELISA as described in Example 6, except that streptavidin was coated onto the ELISA plate at 250ng/well, in place of the anti-cadherin antibody. The nucleotide sequence of an ELISA positive clone insert and DBD junction was checked by DNA sequencing using oligonucleotides M13FOR (SEQ ID No 26) and ORSEQBAK (SEQ ID No 36) (see above). A fifty-fold concentrated stock of C-terminal PCC PDCP particles was prepared and tested in ELISA against streptavidin as described in Example 6 (see Figure 7).

This example shows not only that the peptide can be displayed as a C-terminal fusion on a PDCP, but also that *in vivo* modified peptides can be displayed.

Example 8. Construction of a human scFv PDCP display library

This example describes the generation of a human antibody library of scFvs made from an un-immunised human. The overall strategy for the PCR assembly of scFv fragments is similar to that employed by Marks, J. D. et al. 1991, J. Mol. Biol. 222: 581-597. The antibody gene oligonucleotides used to construct the library are derived from the Marke et al., paper and from sequence data extracted from the Kabat database (Kabat, E. A. et al., Sequences of Proteins of Immunological Interest. 4th edition. U.S. Department of Health and Human Services. 1987). The three linker oligonucleotides are described by Zhou et al. (1994, Nucleic Acids Res., 22: 888-889), all oligonucleotides used are detailed in Table 1.

First, mRNA was isolated from peripheral blood lymphocytes and cDNA prepared for four repertoires of antibody genes IgD, IgM, Igκ and Igλ, using four separate cDNA synthesis primers. VH genes were amplified from IgD and IgM primed cDNA, and VL genes were amplified from Igκ and Igλ primed cDNA. A portion of each set of amplified heavy chain or light chain DNA was then spliced with a separate piece of linker DNA encoding the 15 amino acids (Gly$_4$ Ser)$_3$ (Huston, J. S. et al. 1989, Gene, 77: 61). The 3'-end of the VH PCR products and the 5'-end of the VL PCR products overlap the linker sequence as a result of incorporating linker sequence in the JH, Vκ and Vλ family primer sets (Table 1). Each VH-linker or linker-VL DNA product was then spliced with either VH or VL DNA to produce the primary scFv product in a VH-linker-VL configuration. This scFv product was then amplified and cloned into pDM12 as a SfiI-NotI fragment, electroporated into TG1 and a concentrated PDCP stock prepared.

mRNA isolation and cDNA synthesis.

Human lymphocyte mRNA was purified as described in Example 2. Separate cDNA reactions were performed with IGDCDNAFOR (SEQ ID No 37), IGMCDNAFOR (SEQ ID No 38), IGKCDNAFOR (SEQ ID No 39) and IGλCDNAFOR (SEQ ID No 40) oligonucleotides. 50pmol of each primer was added to approximately 5μg of mRNA in 20μl of nuclease free water and heated to 70°C for 5 minutes and cooled rapidly on ice, then made up to a final reaction volume of 100μl containing 50mM Tris pH 8.3, 75mM KCl, 3mM $MgCl_2$, 10mM DTT, 0.5mM dNTPs, and 2000 units of Superscript II reverse transcriptase (Life Technologies, Paisley, Scotland, U.K.). The reactions were incubated at 37°C for two hours, then heated to 95°C for 5 minutes.

Primary PCRs.

For the primary PCR amplifications separate amplifications were set up for each family specific primer with either an equimolar mixture of the JHFOR primer set (SEQ ID Nos 41 to 44) for IgM and IgD cDNA, or with SCFVκFOR (SEQ ID No 51) or SCFVλFOR (SEQ ID No 52) for IgK or Igλ cDNA respectively e.g. VH1BAK and JHFOR set; Vκ2BAK (SEQ ID No 54) and SCFVκFOR (SEQ ID No 51); Vλ3aBAK (SEQ ID No 66) and SCFVλFOR (SEQ ID No 52) etc. Thus, for IgM, IgD and Igκ cDNA six separate reactions were set up, and seven for Igλ cDNA. A 50μl reaction mixture was prepared containing 2μl cDNA, 25pmol of the appropriate FOR and BAK primers, 0.1mM dNTPs, 2.5 units Taqplus DNA polymerase, and 1x High Salt PCR reaction buffer (20mM Tris-HCl pH 9.2, 60mM KCl, 2mM $MgCl_2$) (Stratagene Ltd, Cambridge, U.K.).

Reactions were amplified on a Techne Progene thermal cycler for 30 cycles of 94°C, 1 minute; 60°C, 1 minute; 72°C, 2 minutes, followed by 10 minutes at 72°C. Fifty microlitres of all 25 reaction products were electrophoresed on an agarose gel, excised and products purified from the gel using a Geneclean II kit according to the manufacturers instructions (Bio101, La Jolla, California, U.S.A.). All sets of IgD, IgM, IgK or Igλ reaction products were pooled to produce VH or VL DNA sets for each of the four repertoires. These were then adjusted to approximately 20ng/μl.

Preparation of linker.

Linker product was prepared from eight 100μl reactions containing 5ng LINKAMP3T (SEQ ID No 76) template oligonucleotide, 50pmol of LINKAMP3 (SEQ ID No 74) and LINKAMP5 (SEQ ID No 75) primers, 0.1mM dNTPs, 2.5 units Taqplus DNA polymerase, and 1x High Salt PCR reaction buffer (20mM Tris-HCl pH 9.2, 60mM KCl, 2mM $MgCl_2$) (Stratagene Ltd, Cambridge, U.K.). Reactions were amplified on a Techne Progene thermal cycler for 30 cycles of 94°C, 1 minute; 60°C, 1 minute; 72°C, 1 minute, followed by 10 minutes at 72°C. All reaction product was electrophoresed on a 2% low melting point agarose gel, excised and products purified from the gel using a Mermaid kit according to the manufacturers instructions (Bio101, La Jolla, California, U.S.A.) and adjusted to 5ng/μl.

First stage linking.

Four linking reactions were prepared for each repertoire using 20ng of VH or VL DNA with 5ng of Linker DNA in 100μl reactions containing (for IgM or IgD VH) 50pmol of LINKAMPFOR and VH1-6BAK set, or, 50pmol LINKAMPBAK and either SCFVκFOR (Igκ) or SCFVλFOR (Igλ), 0.1mM dNTPs, 2.5 units Taq DNA polymerase, and

1x PCR reaction buffer (10mM Tris-HCl pH 9.0, 5mM KCl, 0.01% Triton X*-100, 1.5mM $MgCl_2$) (Promega Ltd, Southampton, U.K.). Reactions were amplified on a Techne Progene thermal cycler for 30 cycles of 94°C, 1 minute; 60°C, 1 minute; 72°C, 2 minutes, followed by 10 minutes at 72°C. Reaction products were electrophoresed on an agarose gel, excised and products purified from the gel using a Geneclean II kit according to the manufacturers instructions (Bio101, La Jolla, California, U.S.A.) and adjusted to 20ng/μl.

Final linking and reamplification.

To prepare the final scFv DNA products, five 100μl reactions were performed for VH-LINKER plus VL DNA, and, five 100μl reactions were performed for VH plus LINKER-VL DNA for each of the four final repertoires (IgM VH-VK, VH-Vλ; IgD VH-VK, VH-Vλ) as described in step (d) above using 20ng of each component DNA as template. Reaction products were electrophoresed on an agarose gel, excised and products purified from the gel using a Geneclean II kit according to the manufacturers instructions (Bio101, La Jolla, California, U.S.A.) and adjusted to 20ng/μl. Each of the four repertoires was then re-amplified in a 100μl reaction volume containing 2ng of each linked product, with 50pmol VHBAK1-6 (SEQ ID Nos 53 to 58) and either the JKFOR (SEQ ID Nos 66 to 70) or JλFOR (SEQ ID Nos 71 to 73) primer sets, in the presence of 0.1mM dNTPs, 2.5 units Taq DNA polymerase, and 1x PCR reaction buffer (10mM Tris-HCl pH 9.0, 5mM KCl, 0.01% Triton X*-100, 1.5mM $MgCl_2$) (Promega Ltd, Southampton, U.K.). Thirty reactions were performed per repertoire to generate enough DNA for cloning. Reactions were amplified on a Techne Progene thermal cycler for 25 cycles of 94°C, 1 minute; 65°C, 1 minute; 72°C, 2 minutes, followed by 10 minutes at 72°C. Reaction products were phenol-chloroform extracted, ethanol precipitated, vacuum dried and re-suspended in 80μl nuclease free water.

Cloning into pDM12.

Each of the four repertoires was SfiI-NotI digested, and electrophoresed on an agarose gel, excised and products purified from the gel using a Geneclean II kit according to the manufacturers instructions (Bio101, La Jolla, California, U.S.A.). Each of the four repertoires was ligated overnight at 16°C in 140μl with 10μg of SfiI-NotI cut pDM12 prepared as in Example 2, and 12 units of T4 DNA ligase (Life Technologies, Paisley, Scotland, U.K.). After incubation the ligations were adjusted to 200μl with nuclease free water, and DNA precipitated with 1μl 20mg/ml glycogen, 100μl 7.5M ammonium acetate and 900μl ice-cold (-20°C) absolute ethanol, vortex mixed and spun at 13,000rpm for 20 minutes in a microfuge to pellet DNA. The pellets were washed with 500μl ice-cold 70% ethanol by centrifugation at 13,000rpm for 2 minutes, then vacuum dried and re-suspended in 10μl DEPC-treated water. 1μl aliquots of each repertoire was electroporated into 80μl *E. coli* (TG1). Cells were grown in 1ml SOC medium per cuvette used for 1 hour at 37°C, and plated onto 2xTY agar plates supplemented with 1% glucose and 100μg/ml ampicillin. 10^{-4}, 10^{-5} and 10^{-6} dilutions of the electroporated bacteria were also plated to assess library size. Colonies were allowed to grow overnight at 30°C. Cloning into SfiI-NotI digested pDM12 yielded an IgM-κ/λ repertoire of 1.16×10^9 clones, and an IgD-κ/λ repertoire of 1.21×10^9 clones.

Preparation of PDCP stock.

Separate PDCP stocks were prepared for each repertoire library. The bacteria were then scraped off the plates into 30ml 2xTY broth supplemented with 20% glycerol, 1%

glucose and 100μg/ml ampicillin. 3ml was added to a 50ml 2xTY culture broth supplemented with 1% glucose and 100μg/ml ampicillin and infected with 10^{11} kanamycin resistance units (kru) M13K07 helper phage at 37°C for 30 minutes without shaking, then for 30 minutes with shaking at 200rpm. Infected bacteria were transferred to 500ml 2xTY broth supplemented with 25μg/ml kanamycin, 100μg/ml ampicillin, and 20μM IPTG, then incubated overnight at 30°C, shaking at 200rpm. Bacteria were pelleted at 4000rpm for 20 minutes in 50ml Falcon tubes, and 80ml 2.5M NaCl/20% PEG 6000 was added to 400ml of particle supernatant, mixed vigorously and incubated on ice for 1 hour to precipitate PDCP particles. Particles were pelleted at 11000rpm for 30 minutes in 250ml Oakridge tubes at 4°C in a Sorvall RC5B centrifuge, then resuspended in 40ml water and 8ml 2.5M NaCl/20% PEG 6000 added to reprecipitate particles, then incubated on ice for 20 minutes. Particles were again pelleted at 11000rpm for 30 minutes in 50ml Oakridge tubes at 4°C in a Sorvall RC5B centrifuge, then resuspended in 5ml PBS buffer, after removing all traces of PEG/NaCl with a pipette. Bacterial debris was removed by a 5 minute 13500rpm spin in a microcentrifuge. The supernatant was filtered through a 0.45μm polysulfone syringe filter, adjusted to 20% glycerol and stored at -70°C.

Example 9. Isolation of binding activity from a N-terminal display PDCP library of human scFvs

The ability to select binding activities to a target of interest from a human antibody library is important due to the possibility of generating therapeutic human antibodies. In addition, such libraries allow the isolation of antibodies to targets which cannot be used for traditional methods of antibody generation due to toxicity, low immunogenicity or ethical considerations. In this example we demonstrate the isolation of specific binding activities against a peptide antigen from a PDCP library of scFvs from an un-immunised human.

The generation of the library, used for the isolation of binding activities in this example, is described in Example 8.

Substance P is an eleven amino acid neuropeptide involved in inflammatory and pain responses *in vivo*. It has also been implicated in a variety of disorders such as psoriasis and asthma amongst others (Misery, L. 1997, Br. J. Dertmatol., 137: 843-850; Maggi, C. A. 1997, Regul. Pept. 70: 75-90; Choi, D. C. & Kwon, O.J., 1998, Curr. Opin. Pulm. Med., 4: 16-24). Human antibodies which neutralise this peptide may therefore have some therapeutic potential. As this peptide is too small to coat efficiently on a tube, as described in Example 3, selection of binding activities was performed in-solution, using N-terminal biotinylated substance P and capturing bound PDCP particles on streptavidin-coated magnetic beads.

Enrichment for substance P binding PDCP particles.
An aliquot of approximately 10^{13} a.r.u. IgM and IgD scFv library stock was mixed with 1μg biotinylated substance P in 800μl 4% BSA/0.1% Tween 20/PBS, and allowed to bind for two hours at ambient temperature. Bound PDCPs were then captured onto 1ml of BSA blocked streptavidin coated magnetic beads for 10 minutes at ambient temperature. The beads were captured to the side of the tube with a magnet (Promega), and unbound material discarded. The beads were washed eight times with 1ml PBS/0.1% Tween 20/ 10μg/ml streptavidin, then two times

with 1ml of PBS by magnetic capture and removal of wash buffer. After the final wash bound PDCPs were eluted with 1ml of freshly prepared 0.1M triethylamine for 10 minutes, the beads were captured, and eluted particles transferred to 0.5ml 1M Tris-HCl pH 7.4. Neutralised particles were added to 10ml log phase TG1 *E. coli* bacteria and incubated at 37°C without shaking for 30 minutes, then with shaking at 200rpm for 30 minutes. 10^{-3}, 10^{-4} & 10^{-5} dilutions of the infected culture were prepared to estimate the number of particles recovered, and the remainder was spun at 4000 rpm for 10 minutes, and the pellet resuspended in 300μl 2xTY medium by vortex mixing. Bacteria were plated onto 2xTY agar plates supplemented with 1% glucose and 100μg/ml ampicillin. Colonies were allowed to grow overnight at 30°C. A 100-fold concentrated PDCP stock was prepared from a 200ml amplified culture of these bacteria as described above, and 0.5ml used in as second round of selection with 500ng biotinylated substance P. For this round 100μg/ml streptavidin was included in the wash buffer.

ELISA identification of binding clones.

Binding clones were identified by ELISA of 96 individual PDCP cultures prepared as described in Example 3 from colonies recovered after the second round of selection. A Dynatech Immulon 4 ELISA plate was coated with 200ng/well streptavidin in 100μl /well PBS for 1 hour at 37°C. The plate was washed 3x200μl/well PBS and incubated with 10ng/well biotinylated substance P in 100μl /well PBS for 30 minutes at 37°C The plate was washed 3x200μl/well PBS and blocked for 1 hour at 37°C with 200μl/well 2% Marvel non-fat milk powder/PBS and then washed 2x200μl/well PBS. 50μl PDCP culture supernatant was added to each well containing 50μl/well 4% Marvel/PBS, and allowed to bind for 1 hour at ambient temperature. The plate was washed three times with 200μl/well PBS/0.1% Tween 20, then three times with 200μl/well PBS. Bound PDCPs were detected with 100μl/well, 1:5000 diluted anti-M13-HRP conjugate (Pharmacia) in 2% Marvel/PBS for 1 hour at ambient temperature and the plate washed six times as above. The plate was developed for 10 minutes at ambient temperature with 100μl/well freshly prepared TMB (3,3',5,5'-Tetramethylbenzidine) substrate buffer (0.005% H_2O_2, 0.1mg/ml TMB in 24mM citric acid/52mM sodium phosphate buffer pH 5.2). The reaction was stopped with 100μl/well 12.5% H_2SO_4 and read at 450nm. Out of 96 clones tested, 10 gave signals greater than twice background (background = 0.05).

Characterization of a binding clone.

A 50-fold concentrated PDCP stock was prepared from a 100ml amplified culture of a single ELISA positive clone as described above. 10μl per well of this stock was tested in ELISA as described above for binding to streptavidin, streptavidin-biotinylated-substance P and streptavidin-biotinylated-CGRP (N-terminal biotinylated). Binding was only observed in streptavidin-biotinylated-substance P coated wells indicating that binding was specific. In addition, binding to streptavidin-biotinylated substance P was completely inhibited by incubating the PDCP with 1μg/ml free substance P (see Figure 8). The scFv VH (SEQ ID Nos 15 and 16) and VL (SEQ ID Nos 17 and 18) DNA and amino acid sequence was determined by DNA sequencing with oligonucleotides M13REV (SEQ ID No27) and ORSEQFOR (SEQ ID No 36) and is shown in Figure 9.

The results indicate that target binding activities can be isolated from PDCP display libraries of human scFv fragments.

Example 10

In another example the invention provides methods for screening a DNA library whose members require more than one chain for activity, as required by, for example, antibody Fab fragments for ligand binding. To increase the affinity of an antibody of known heavy and light chain sequence, libraries of unknown light chains co-expressed with a known heavy chain are screened for higher affinity antibodies. The known heavy chain antibody DNA sequence is joined to a nucleotide sequence encoding a oestrogen receptor DNA binding domain in a phage vector which does not contain the oestrogen receptor HRE sequence. The antibody DNA sequence for the known heavy chain (VH and CH1) gene is inserted in the 5′ region of the oestrogen receptor DBD DNA, behind an appropriate promoter and translation sequences and a sequence encoding a signal peptide leader directing transport of the downstream fusion protein to the periplasmic space. The library of unknown light chains (VL and CL) is expressed separately from a phagemid expression vector which also contains the oestrogen receptor HRE sequence. Thus when both heavy and light chains are expressed in the same host cell, following infection with the phage containing the heavy chain-DBD fusion, the light chain phagemid vector is preferentially packaged into mature phage particles as single stranded DNA, which is bound by the heavy chain-DBD fusion protein during the packaging process. The light chain proteins are transported to the periplasm where they assemble with the heavy chain that is fused to the DBD protein as it exits the cell on the PDCP. In this example the DBD fusion protein and the HRE DNA sequences are not encoded on the same vector, the unknown peptide sequences are present on the same vector as the HRE sequence. Peptide display carrier packages (PDCP) which encode the protein of interest can then be selected by means of a ligand specific for the antibody.

Table 1 (i) Oligonucleotide primers used for human scFv library construction

cDNA synthesis primers

IgMCDNAFOR	TGGAAGAGGCACGTTCTTTTCTTT
IgDCDNAFOR	CTCCTTCTTACTCTTGCTGGCGGT
IgκCDNAFOR	AGACTCTCCCCTGTTGAAGCTCTT
IgλCDNAFOR	TGAAGATTCTGTAGGGGCCACTGTCTT

JHFOR primers

JH1-2FOR	TGAACCGCCTCCACCTGAGGAGACGGTGACCAGGGTGCC
JH3FOR	TGAACCGCCTCCACCTGAAGAGACGGTGACCATTGTCCC
JH4-5FOR	TGAACCGCCTCCACCTGAGGAGACGGTGACCAGGGTTCC
JH6FOR	TGAACCGCCTCCACCTGAGGAGACGGTGACCGTGGTCCC

VH familyBAKprimers

VH1BAK	TTTTTGGCCCAGCCGGCCATGGCCCAGGTGCAGCTGGTGCAGTCTGG
VH2BAK	TTTTTGGCCCAGCCGGCCATGGCCCAGGTCAACTTAAGGGAGTCTGG
VH3BAK	TTTTTGGCCCAGCCGGCCATGGCCGAGGTGCAGCTGGTGGAGTCTGG
VH4BAK	TTTTTGGCCCAGCCGGCCATGGCCCAGGTGCAGCTGCAGGAGTCGGG
VH5BAK	TTTTTGGCCCAGCCGGCCATGGCCGAGGTGCAGCTGTTGCAGTCTGC
VH6BAK	TTTTTGGCCCAGCCGGCCATGGCCCAGGTACAGCTGCAGCAGTCAGG

Light chain FOR primers

SCFVKFOR	TTATTCGCGGCCGCCTAAACAGAGGCAGTTCCAGATTTC
SCFVλFOR	GTCACTTGCGGCCGCCTACAGTGTGGCCTTGTTGGCTTG

VK family BAK primers

VK1BAK	TCTGGCGGTGGCGGATCGGACATCCAGATGACCCAGTCTCC
VK2BAK	TCTGGCGGTGGCGGATCGGATGTTGTGATGACTCAGTCTCC
VK3BAK	TCTGGCGGTGGCGGATCGGAAATTGTGTTGACGCAGTCTCC
VK4BAK	TCTGGCGGTGGCGGATCGGACATCGTGATGACCCAGTCTCC
VK5BAK	TCTGGCGGTGGCGGATCGGAAACGACACTCACGCAGTCTCC
VK6BAK	TCTGGCGGTGGCGGATCGGAAATTGTGCTGACTCAGTCTCC

JK FOR primers

JK1FOR	TTCTCGTGCGGCCGCCTAACGTTTGATTTCCACCTTGGTCCC
JK2FOR	TTCTCGTGCGGCCGCCTAACGTTTGATCTCCAGCTTGGTCCC
JK3FOR	TTCTCGTGCGGCCGCCTAACGTTTGATATCCACTTTGGTCCC
JK4FOR	TTCTCGTGCGGCCGCCTAACGTTTGATCTCCACCTTGGTCCC
JK5FOR	TTCTCGTGCGGCCGCCTAACGTTTAATCTCCAGTCGTGTCCC

Vλ family BAK primers

Vλ1BAK	TCTGGCGGTGGCGGATCGCAGTCTGTGTTGACGCAGCCGCC
Vλ2BAK	TCTGGCGGTGGCGGATCGCAGTCTGCCCTGACTCAGCCTGC

Table 1 (ii) Oligonucleotide primers used for human scFv library construction

Vλ3aBAK	TCTGGCGGTGGCGGATCGTCCTATGTGCTGACTCAGCCACC
Vλ3bBAK	TCTGGCGGTGGCGGATCGTCTTCTGAGCTGACTCAGGACCC
Vλ4BAK	TCTGGCGGTGGCGGATCGCACGTTATACTGACTCAACCGCC
Vλ5BAK	TCTGGCGGTGGCGGATCGCAGGCTGTGCTCACTCAGCCGTC
Vλ6BAK	TCTGGCGGTGGCGGATCGAATTTTATGCTGACTCAGCCCCA

Jλ primers

Jλ1FOR	TTCTCGTGCGGCCGCCTAACCTAGGACGGTGACCTTGGTCCC
Jλ2-3FOR	TTCTCGTGCGGCCGCCTAACCTAGGACGGTCAGCTTGGTCCC
Jλ4-5FOR	TTCTCGTGCGGCCGCCTAACCTAAAACGGTGAGCTGGGTCCC

Linker primers

LINKAMP3	CGATCCGCCACCGCCAGA
LINKAMP5	GTCTCCTCAGGTGGAGGC
LINKAMP3T	CGATCCGCCACCGCCAGAGCCACCTCCGCCTGAACCGCCTCCACCTGAGGAGAC

Claims

1. A peptide display carrier package (PDCP), said package comprising a recombinant polynucleotide-chimeric protein complex wherein the chimeric protein has a nucleotide binding portion and a target peptide portion, wherein said recombinant polynucleotide comprises a nucleotide sequence motif which is specifically bound by said nucleotide binding portion, and wherein at least the chimeric protein-encoding portion of the recombinant polynucleotide not bound by the chimeric protein nucleotide binding portion is protected by a binding moeity.

2. A peptide display carrier package (PDCP) as claimed in Claim 1, wherein said chimeric protein-encoding portion of the recombinant polynucleotide not bound by the chimeric protein nucleotide binding portion is protected by a non-sequence-specific protein.

3. A peptide display carrier package (PDCP) as claimed in Claim 2, wherein said non-sequence-specific protein is a viral coat protein.

4. A peptide display carrier package (PDCP) as claimed in any one of Claims 1 to 3, wherein said target peptide portion is displayed externally on the package.

5. A peptide display carrier package (PDCP) as claimed in any one of Claims 1 to 4 wherein said recombinant polynucleotide includes a linker sequence between the nucleotide sequence encoding the nucleotide binding portion and the nucleotide sequence encoding the target peptide portion.

6. A peptide display carrier package (PDCP) as claimed in any one of Claims 1 to 5 wherein said recombinant polynucleotide has two or more nucleotide sequence motifs each of which can be bound by the nucleotide binding portion of the chimeric protein.

7. A peptide display carrier package (PDCP) as claimed in any one of Claims 1 to 6 wherein said nucleotide binding portion is a DNA binding domain of an oestrogen or progesterone receptor.

8. A peptide display carrier package (PDCP) as claimed in any one of Claims 1 to 7 wherein said recombinant polynucleotide is bound to said chimeric protein as single stranded DNA.

9. A peptide display carrier package (PDCP) as claimed in any one of Claims 1 to 8 wherein said target peptide portion is located at the N and/or C terminal of the chimeric protein.

10. A peptide display carrier package (PDCP) as claimed in any one of Claims 1 to 9 which is produced in a host cell transformed with said recombinant polynucleotide and extruded therefrom without lysis of the host cell.

11. A recombinant polynucleotide comprising a nucleotide sequence encoding a chimeric protein having a nucleotide binding portion operably linked to a target peptide portion, wherein said polynucleotide includes a specific nucleotide sequence motif which is bound by the nucleotide binding portion of said chimeric protein and further encoding a non-sequence-specific nucleotide

binding protein.

12. A recombinant polynucleotide as claimed in Claim 11 wherein said non-sequence-specific nucleotide binding protein is a viral coat protein.

13. A recombinant polynucleotide as claimed in either one of Claims 11 and 12 which includes a linker sequence between the nucleotide sequence encoding the nucleotide binding portion and the nucleotide sequence encoding the target peptide portion.

14. A recombinant polynucleotide as claimed in any one of Claims 11 to 13 which has two or more nucleotide sequence motifs each of which can be bound by the nucleotide binding portion of the chimeric protein.

15. A recombinant polynucleotide as claimed in any one of Claims 11 to 14 wherein said nucleotide binding portion is a DNA binding domain of an oestrogen or progesterone receptor.

16. A recombinant polynucleotide as claimed in any one of Claims 11 to 15 wherein said recombinant polynucleotide is bound to said chimeric protein as single stranded DNA.

17. A genetic construct or set of genetic constructs which collectively comprises a polynucleotide having a sequence which includes:

i) a sequence encoding a nucleotide binding portion able to recognise and bind to a specific sequence motif;

ii) the sequence motif recognised and bound by the nucleotide binding portion encoded by (i);

iii) a restriction enzyme site which permits insertion of a polynucleotide, said site being designed to operably link said polynucleotide to the sequence encoding the nucleotide binding portion so that expression of the operably linked polynucleotide sequences yields a chimeric protein; and

iv) a sequence encoding a nucleotide binding protein which binds non-specifically to naked polynucleotide.

18. A genetic construct or set of genetic constructs as claimed in Claim 17 wherein a linker sequence is located between the nucleotide sequence encoding the nucleotide binding portion and the site for insertion of the polynucleotide.

19. A genetic construct or set of genetic constructs as claimed in either one of Claims 17 and 18 which includes a vector pDM12, pDM14 or pDM16, deposited at NCIMB under Nos 40970, 40971 and 40972 respectively.

20. A method of constructing a genetic library, said method comprising:

a) constructing multiple copies of a recombinant vector comprising a polynucleotide sequence which encodes a nucleotide binding portion able to recognise and bind to a specific sequence motif;

b) operably linking each said vector to a polynucleotide encoding a target polypeptide, such that expression of said operably linked vector results in expression of a chimeric protein comprising said target peptide and said nucleotide binding portions; wherein said multiple copies of said operably linked vectors collectively express a library of target peptide portions;

c) transforming host cells with the vectors of step b);

d) culturing the host cells of step c) under conditions suitable for expression of said chimeric protein;

e) providing a recombinant polynucleotide comprising the nucleotide sequence motif specifically recognised by the nucleotide binding portion and exposing this polynucleotide to the chimeric protein of step d) to yield a polynucleotide-chimeric protein complex; and

f) causing production of a non-sequence-specific moiety able to bind to the non-protected portion of the polynucleotide encoding the chimeric protein to form a peptide display carrier package.

21. A method of screening a genetic library, said method comprising:

a) exposing the polynucleotide members of said library to multiple copies of a genetic construct comprising a nucleotide sequence encoding a nucleotide binding portion able to recognise and bind to a specific sequence motif, under conditions suitable for the polynucleotides of said library each to be individually ligated into one copy of said genetic construct, to create a library of recombinant polynucleotides;

b) exposing said recombinant polynucleotides to a population of host cells, under conditions suitable for transformation of said host cells by said recombinant polynucleotides;

c) selecting for transformed host cells;

d) exposing said transformed host cells to conditions suitable for expression of said recombinant polynucleotide to yield a chimeric protein; and

e) providing a recombinant polynucleotide comprising the nucleotide sequence motif specifically recognised by the nucleotide binding portion and exposing this polynucleotide to the chimeric protein of step d) to yield a polynucleotide-chimeric protein complex;

f) protecting any exposed portions of the polynucleotide in the complex of step e) to form a peptide display carrier package; and

g) screening said peptide display carrier package to select only those packages displaying a target peptide portion having the characteristics required.

22. A method as claimed in Claim 21 wherein the peptide display package carrier is extruded from the host cell without lysis thereof.

23. A polynucleotide comprising a nucleotide sequence substantially as set out in SEQ ID No. 15 or SEQ ID No. 17.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: Rowett Research Services Limited
(B) STREET: Greenburn
(C) CITY: Buckburn ABERDEEN
(E) COUNTRY: United Kingdom
(F) POSTAL CODE (ZIP): AB21 9SB

(ii) TITLE OF INVENTION: Chimeric binding peptide library screening method

(iii) NUMBER OF SEQUENCES: 76

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 38 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

TCAGGTCAGA GTGACCTGAG CTAAAATAAC ACATTCAG 38

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 38 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

```
AGTCCAGTCT CACTGGACTC GATTTTATTG TGTAAGTC                              38
```

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 521 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:41..475

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

```
AAGCTTGCAT GCAAATTCTA TTTCAAGGAG ACAGTCATAA ATG AAA TAC CTA TTG      55
                                           Met Lys Tyr Leu Leu
                                             1               5

CCT ACG GCA GCC GCT GGA TTG TTA TTA CTC GCG GCC CAG CCG GCC ATG     103
Pro Thr Ala Ala Ala Gly Leu Leu Leu Leu Ala Ala Gln Pro Ala Met
                 10                  15                  20

GCC CAA GTG CAG CTG CAG TAA TAG GCG GCC GCA GGG GGA GGA GGG TCC     151
Ala Gln Val Gln Leu Gln  *   *  Ala Ala Ala Gly Gly Gly Gly Ser
             25                  30                  35

ATG GAA TCT GCC AAG GAG ACT CGC TAC TGT GCA GTG TGC AAT GAC TAT     199
Met Glu Ser Ala Lys Glu Thr Arg Tyr Cys Ala Val Cys Asn Asp Tyr
         40                  45                  50

GCT TCA GGC TAC CAT TAT GGA GTC TGG TCC TGT GAG GGC TGC AAG GCC     247
Ala Ser Gly Tyr His Tyr Gly Val Trp Ser Cys Glu Gly Cys Lys Ala
     55                  60                  65

TTC TTC AAG AGA AGT ATT CAA GGA CAT AAC GAC TAT ATG TGT CCA GCC     295
Phe Phe Lys Arg Ser Ile Gln Gly His Asn Asp Tyr Met Cys Pro Ala
 70                  75                  80                  85

ACC AAC CAG TGC ACC ATT GAT AAA AAC AGG AGG AAG AGC TGC CAG GCC     343
Thr Asn Gln Cys Thr Ile Asp Lys Asn Arg Arg Lys Ser Cys Gln Ala
                 90                  95                 100

TGC CGG CTC CGT AAA TGC TAC GAA GTG GGA ATG ATG AAA GGT GGG ATA     391
Cys Arg Leu Arg Lys Cys Tyr Glu Val Gly Met Met Lys Gly Gly Ile
            105                 110                 115

CGA AAA GAC CGA AGA GGA GGG AGA ATG TTG AAA CAC AAG CGC CAG AGA     439
Arg Lys Asp Arg Arg Gly Gly Arg Met Leu Lys His Lys Arg Gln Arg
```

```
        120                     125                     130

GAT GAT GGG GAG GGC AGG GGT GAA GTG GGG TCT TGA TAATCAGGTC                    485
Asp Asp Gly Glu Gly Arg Gly Glu Val Gly Ser  *
    135                 140                 145

AGAGTGACCT GAGCTAAAAT AACACATTCA GAATTC                                       521
```

(2) INFORMATION FOR SEQ ID NO: 4:

```
       (i) SEQUENCE CHARACTERISTICS:
          (A) LENGTH: 145 amino acids
          (B) TYPE: amino acid
          (D) TOPOLOGY: linear

     (ii) MOLECULE TYPE: protein
     (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Met Lys Tyr Leu Leu Pro Thr Ala Ala Ala Gly Leu Leu Leu Leu Ala
  1               5                  10                  15

Ala Gln Pro Ala Met Ala Gln Val Gln Leu Gln  *   *  Ala Ala Ala
             20                  25                  30

Gly Gly Gly Gly Ser Met Glu Ser Ala Lys Glu Thr Arg Tyr Cys Ala
         35                  40                  45

Val Cys Asn Asp Tyr Ala Ser Gly Tyr His Tyr Gly Val Trp Ser Cys
     50                  55                  60

Glu Gly Cys Lys Ala Phe Phe Lys Arg Ser Ile Gln Gly His Asn Asp
 65                  70                  75                  80

Tyr Met Cys Pro Ala Thr Asn Gln Cys Thr Ile Asp Lys Asn Arg Arg
                 85                  90                  95

Lys Ser Cys Gln Ala Cys Arg Leu Arg Lys Cys Tyr Glu Val Gly Met
            100                 105                 110

Met Lys Gly Gly Ile Arg Lys Asp Arg Arg Gly Gly Arg Met Leu Lys
        115                 120                 125

His Lys Arg Gln Arg Asp Asp Gly Glu Gly Arg Gly Glu Val Gly Ser
    130                 135                 140

 *
145
```

(2) INFORMATION FOR SEQ ID NO: 5:

```
      (i) SEQUENCE CHARACTERISTICS:
          (A) LENGTH: 102 base pairs
          (B) TYPE: nucleic acid
          (C) STRANDEDNESS: double
          (D) TOPOLOGY: linear

     (ii) MOLECULE TYPE: cDNA

    (iii) HYPOTHETICAL: NO

     (iv) ANTI-SENSE: NO
```

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:1..102

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

```
AAA CGA ACT GTG GCT GCA CCA TCT GTC TTC ATC TTC CCG CCA TCT GAT        48
Lys Arg Thr Val Ala Ala Pro Ser Val Phe Ile Phe Pro Pro Ser Asp
                150                 155                 160

GAG CAG TTG AAA TCT GGA ACT GCC TCT GTT GTG TGC CTG CTG AAT AAC        96
Glu Gln Leu Lys Ser Gly Thr Ala Ser Val Val Cys Leu Leu Asn Asn
            165                 170                 175

TTC TAT                                                               102
Phe Tyr
```

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 34 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

```
Lys Arg Thr Val Ala Ala Pro Ser Val Phe Ile Phe Pro Pro Ser Asp
  1               5                  10                  15

Glu Gln Leu Lys Ser Gly Thr Ala Ser Val Val Cys Leu Leu Asn Asn
            20                  25                  30

Phe Tyr
```

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 150 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:1..150

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

```
ATG GCC CAG CCC ACC ACG CGT CCG GGC CAA GGG ACA CGA CTG GAC ATT          48
Met Ala Gln Pro Thr Thr Arg Pro Gly Gln Gly Thr Arg Leu Asp Ile
 35                  40                  45                  50

AAA CGA ACT GTG GCT GCA CCA TCT GTC TTC ATC TTC CCG CCA TCT GAT          96
Lys Arg Thr Val Ala Ala Pro Ser Val Phe Ile Phe Pro Pro Ser Asp
                 55                  60                  65

GAG CAG TTG AAA TCT GGA ACT GCC TCT GTT GTG TGC CTG CTG AAT AAC         144
Glu Gln Leu Lys Ser Gly Thr Ala Ser Val Val Cys Leu Leu Asn Asn
             70                  75                  80

TTC TAT                                                                 150
Phe Tyr
```

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 50 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

```
Met Ala Gln Pro Thr Thr Arg Pro Gly Gln Gly Thr Arg Leu Asp Ile
  1               5                  10                  15

Lys Arg Thr Val Ala Ala Pro Ser Val Phe Ile Phe Pro Pro Ser Asp
             20                  25                  30

Glu Gln Leu Lys Ser Gly Thr Ala Ser Val Val Cys Leu Leu Asn Asn
         35                  40                  45

Phe Tyr
     50
```

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 150 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:1..150

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

```
ATG GCC CAG TCC CAC CAC GCG TCC GGC GGA GGG ACC AAG GTG GAG ATC      48
Met Ala Gln Ser His His Ala Ser Gly Gly Gly Thr Lys Val Glu Ile
                55                  60                  65

AAA CGA ACT GTG GCT GCA CCA TCT GTC TTC ATC TTC CCG CCA TCT GAT      96
Lys Arg Thr Val Ala Ala Pro Ser Val Phe Ile Phe Pro Pro Ser Asp
            70                  75                  80

GAG CAG TTG AAA TCT GGA ACT GCC TCT GTT GTG TGC CTG CTG AAT AAC     144
Glu Gln Leu Lys Ser Gly Thr Ala Ser Val Val Cys Leu Leu Asn Asn
        85                  90                  95

TTC TAT                                                             150
Phe Tyr
    100
```

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 50 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

```
Met Ala Gln Ser His His Ala Ser Gly Gly Gly Thr Lys Val Glu Ile
1               5                   10                  15

Lys Arg Thr Val Ala Ala Pro Ser Val Phe Ile Phe Pro Pro Ser Asp
            20                  25                  30

Glu Gln Leu Lys Ser Gly Thr Ala Ser Val Val Cys Leu Leu Asn Asn
        35                  40                  45

Phe Tyr
    50
```

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 566 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:41..475

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

```
AAGCTTGCAT GCAAATTCTA TTTCAAGGAG ACAGTCATAA ATG AAA TAC CTA TTG      55
```

```
                                             Met Lys Tyr Leu Leu
                                                             55

CCT ACG GCA GCC GCT GGA TTG TTA TTA CTC GCG GCC CAG CCG GCC ATG     103
Pro Thr Ala Ala Ala Gly Leu Leu Leu Leu Ala Ala Gln Pro Ala Met
                 60                  65                  70

GCC GAG GTG CAA CTG CAG TAA TAG GCG GCC GCA GGG GGA GGA GGG TCC     151
Ala Glu Val Gln Leu Gln  *   *  Ala Ala Ala Gly Gly Gly Gly Ser
             75                  80                  85

ATG GAA TCT GCC AAG GAG ACT CGC TAC TGT GCA GTG TGC AAT GAC TAT     199
Met Glu Ser Ala Lys Glu Thr Arg Tyr Cys Ala Val Cys Asn Asp Tyr
         90                  95                 100

GCT TCA GGC TAC CAT TAT GGA GTC TGG TCC TGT GAG GGC TGC AAG GCC     247
Ala Ser Gly Tyr His Tyr Gly Val Trp Ser Cys Glu Gly Cys Lys Ala
    105                 110                 115

TTC TTC AAG AGA AGT ATT CAA GGA CAT AAC GAC TAT ATG TGT CCA GCC     295
Phe Phe Lys Arg Ser Ile Gln Gly His Asn Asp Tyr Met Cys Pro Ala
120                 125                 130                 135

ACC AAC CAG TGC ACC ATT GAT AAA AAC AGG AGG AAG AGC TGC CAG GCC     343
Thr Asn Gln Cys Thr Ile Asp Lys Asn Arg Arg Lys Ser Cys Gln Ala
                140                 145                 150

TGC CGG CTC CGT AAA TGC TAC GAA GTG GGA ATG ATG AAA GGT GGG ATA     391
Cys Arg Leu Arg Lys Cys Tyr Glu Val Gly Met Met Lys Gly Gly Ile
            155                 160                 165

CGA AAA GAC CGA AGA GGA GGG AGA ATG TTG AAA CAC AAG CGC CAG AGA     439
Arg Lys Asp Arg Arg Gly Gly Arg Met Leu Lys His Lys Arg Gln Arg
        170                 175                 180

GAT GAT GGG GAG GGC AGG GGT GAA GTG GGG TCT TGA TAATCAGGTC          485
Asp Asp Gly Glu Gly Arg Gly Glu Val Gly Ser  *
    185                 190                 195

AGAGTGACCT GAGCTAAAAT AACACATTCA GGTCGACTTG GGTCAGTCTG ACCGGGACAA   545

AGTTAATGTA ACCTCGAATT C                                             566
```

(2) INFORMATION FOR SEQ ID NO: 12:

```
        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 145 amino acids
            (B) TYPE: amino acid
            (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: protein
    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

Met Lys Tyr Leu Leu Pro Thr Ala Ala Ala Gly Leu Leu Leu Leu Ala
  1               5                  10                  15

Ala Gln Pro Ala Met Ala Glu Val Gln Leu Gln  *   *  Ala Ala Ala
             20                  25                  30

Gly Gly Gly Gly Ser Met Glu Ser Ala Lys Glu Thr Arg Tyr Cys Ala
         35                  40                  45
```

```
Val Cys Asn Asp Tyr Ala Ser Gly Tyr His Tyr Gly Val Trp Ser Cys
    50                  55                  60

Glu Gly Cys Lys Ala Phe Phe Lys Arg Ser Ile Gln Gly His Asn Asp
65                  70                  75                  80

Tyr Met Cys Pro Ala Thr Asn Gln Cys Thr Ile Asp Lys Asn Arg Arg
                85                  90                  95

Lys Ser Cys Gln Ala Cys Arg Leu Arg Lys Cys Tyr Glu Val Gly Met
            100                 105                 110

Met Lys Gly Gly Ile Arg Lys Asp Arg Arg Gly Gly Arg Met Leu Lys
        115                 120                 125

His Lys Arg Gln Arg Asp Asp Gly Glu Gly Arg Gly Glu Val Gly Ser
    130                 135                 140

*
145
```

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 539 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:41..481

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

```
AAGCTTGCAT GCAAATTCTA TTTCAAGGAG ACAGTCATAA ATG AAA TAC CTA TTG       55
                                            Met Lys Tyr Leu Leu
                                                            150

CCT ACG GCA GCC GCT GGA TTG TTA TTA CTC GCG GCC CAG CCG GCA ATG      103
Pro Thr Ala Ala Ala Gly Leu Leu Leu Leu Ala Ala Gln Pro Ala Met
                155                 160                 165

GCC GAG ATG GAA TCT GCC AAG GAG ACT CGC TAC TGT GCA GTG TGC AAT      151
Ala Glu Met Glu Ser Ala Lys Glu Thr Arg Tyr Cys Ala Val Cys Asn
            170                 175                 180

GAC TAT GCT TCA GGC TAC CAT TAT GGA GTC TGG TCC TGT GAG GGC TGC      199
Asp Tyr Ala Ser Gly Tyr His Tyr Gly Val Trp Ser Cys Glu Gly Cys
        185                 190                 195

AAG GCC TTC TTC AAG AGA AGT ATT CAA GGA CAT AAC GAC TAT ATG TGT      247
Lys Ala Phe Phe Lys Arg Ser Ile Gln Gly His Asn Asp Tyr Met Cys
```

```
    200                        205                        210

CCA GCC ACC AAC CAG TGC ACC ATT GAT AAA AAC AGG AGG AAG AGC TGC          295
Pro Ala Thr Asn Gln Cys Thr Ile Asp Lys Asn Arg Arg Lys Ser Cys
215                     220                     225                 230

CAG GCC TGC CGG CTC CGT AAA TGC TAC GAA GTG GGA ATG ATG AAA GGT          343
Gln Ala Cys Arg Leu Arg Lys Cys Tyr Glu Val Gly Met Met Lys Gly
                    235                     240                 245

GGG ATA CGA AAA GAC CGA AGA GGA GGG AGA ATG TTG AAA CAC AAG CGC          391
Gly Ile Arg Lys Asp Arg Arg Gly Gly Arg Met Leu Lys His Lys Arg
                250                     255                     260

CAG AGA GAT GAT GGG GAG GGC AGG GGT GAA GTG GGG TCT GGG GGA GGA          439
Gln Arg Asp Asp Gly Glu Gly Arg Gly Glu Val Gly Ser Gly Gly Gly
            265                     270                     275

GGG TCG GCC CAG CCG GCC CTC CTG CAG CTG GCG GCC GCA TAA                  481
Gly Ser Ala Gln Pro Ala Leu Leu Gln Leu Ala Ala Ala  *
    280                     285                     290

CTGATTGAGT CGACTTGGGT CAGTCTGACC GGGACAAAGT TAATGTAACC TCGAATTC          539
```

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 147 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

```
Met Lys Tyr Leu Leu Pro Thr Ala Ala Ala Gly Leu Leu Leu Leu Ala
  1             5                  10                  15

Ala Gln Pro Ala Met Ala Glu Met Glu Ser Ala Lys Glu Thr Arg Tyr
             20                  25                  30

Cys Ala Val Cys Asn Asp Tyr Ala Ser Gly Tyr His Tyr Gly Val Trp
         35                  40                  45

Ser Cys Glu Gly Cys Lys Ala Phe Phe Lys Arg Ser Ile Gln Gly His
     50                  55                  60

Asn Asp Tyr Met Cys Pro Ala Thr Asn Gln Cys Thr Ile Asp Lys Asn
 65                  70                  75                  80

Arg Arg Lys Ser Cys Gln Ala Cys Arg Leu Arg Lys Cys Tyr Glu Val
                 85                  90                  95

Gly Met Met Lys Gly Gly Ile Arg Lys Asp Arg Arg Gly Gly Arg Met
            100                 105                 110

Leu Lys His Lys Arg Gln Arg Asp Asp Gly Glu Gly Arg Gly Glu Val
        115                 120                 125

Gly Ser Gly Gly Gly Gly Ser Ala Gln Pro Ala Leu Leu Gln Leu Ala
    130                 135                 140
```

```
Ala Ala
145
```

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 372 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:1..372

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

```
CAG GTA CAG CTG CAG CAG TCA GGG GGA GGC GTG GTC CAG CCT GGG AGG       48
Gln Val Gln Leu Gln Gln Ser Gly Gly Gly Val Val Gln Pro Gly Arg
        150                 155                 160

TCC CTG AGA CTC TCC TGT GCA GCC TCG GGA TTC CCC TTT AGT ACT TAT       96
Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Pro Phe Ser Thr Tyr
    165                 170                 175

GGC ATG CAC TGG CGC CAG GCT GTC CCA GGC AAG GGG CTG GAG TGG GTG      144
Gly Met His Trp Arg Gln Ala Val Pro Gly Lys Gly Leu Glu Trp Val
180                 185                 190                 195

GCA GTT ATA TCA TAT GAT GGA AGT AAT AAA TAC TAC GCA GAC TCC GTG      192
Ala Val Ile Ser Tyr Asp Gly Ser Asn Lys Tyr Tyr Ala Asp Ser Val
                200                 205                 210

AAG GGC CGA TTC ACC ATC TCC AGA GAC AAT TCC AAG AAC ACG TTG TAT      240
Lys Gly Arg Phe Thr Ile Ser Arg Asp Asn Ser Lys Asn Thr Leu Tyr
            215                 220                 225

CTG CAA ATG AAC AGC CTG AGA GCT GAG GAC ACG GCT GTG TAT TAC TGT      288
Leu Gln Met Asn Ser Leu Arg Ala Glu Asp Thr Ala Val Tyr Tyr Cys
        230                 235                 240

GCG AGA GAT TTA GAC CCC ACC AGG TAT AGC AGT GGC TGG GAC ACT GAC      336
Ala Arg Asp Leu Asp Pro Thr Arg Tyr Ser Ser Gly Trp Asp Thr Asp
    245                 250                 255

TAC TGG GGC CAG GGG CAC CTG GTC ACT GTC TCC TCA                      372
Tyr Trp Gly Gln Gly His Leu Val Thr Val Ser Ser
260                 265                 270
```

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 124 amino acids

(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

```
Gln Val Gln Leu Gln Gln Ser Gly Gly Gly Val Val Gln Pro Gly Arg
 1               5                  10                  15

Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Pro Phe Ser Thr Tyr
            20                  25                  30

Gly Met His Trp Arg Gln Ala Val Pro Gly Lys Gly Leu Glu Trp Val
        35                  40                  45

Ala Val Ile Ser Tyr Asp Gly Ser Asn Lys Tyr Tyr Ala Asp Ser Val
    50                  55                  60

Lys Gly Arg Phe Thr Ile Ser Arg Asp Asn Ser Lys Asn Thr Leu Tyr
65                  70                  75                  80

Leu Gln Met Asn Ser Leu Arg Ala Glu Asp Thr Ala Val Tyr Tyr Cys
                85                  90                  95

Ala Arg Asp Leu Asp Pro Thr Arg Tyr Ser Ser Gly Trp Asp Thr Asp
            100                 105                 110

Tyr Trp Gly Gln Gly His Leu Val Thr Val Ser Ser
        115                 120
```

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 327 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:1..327

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

```
GAA ACG ACA CTC ACG CAG TCT CCA GGC ACC CTG TCT TTG TCT CCG GGG       48
Glu Thr Thr Leu Thr Gln Ser Pro Gly Thr Leu Ser Leu Ser Pro Gly
125                 130                 135                 140

GAA AGA GCC ACC CTC TCC TGC AGG GCC AGT CAG AAT ATT GGC AGC AGC       96
Glu Arg Ala Thr Leu Ser Cys Arg Ala Ser Gln Asn Ile Gly Ser Ser
                145                 150                 155

TCC TTA GCC TGG TAC CAA CAG AAA CCT GGC CAG GCT CCC AGG CTC CTC      144
Ser Leu Ala Trp Tyr Gln Gln Lys Pro Gly Gln Ala Pro Arg Leu Leu
```

```
            160                 165                 170

ATC TAT GGT GCA TCC ACC AGG GCC ACT GGT ATC CCA GCC AGG TTC AGT      192
Ile Tyr Gly Ala Ser Thr Arg Ala Thr Gly Ile Pro Ala Arg Phe Ser
        175                 180                 185

GGC AGT GGG TCA GGG ACA CAA TTC ACT CTC ACC ATC AGC AGC CTG CAG      240
Gly Ser Gly Ser Gly Thr Gln Phe Thr Leu Thr Ile Ser Ser Leu Gln
    190                 195                 200

TCT GAA GAT TTT GCA GTT TAT TAC TGT CAG CAG TAT AAT TTC TGG CCA      288
Ser Glu Asp Phe Ala Val Tyr Tyr Cys Gln Gln Tyr Asn Phe Trp Pro
205                 210                 215                 220

TTC ACT TTT GGC CCT GGG ACC AAG CTG GAG ATC AAA CGT                  327
Phe Thr Phe Gly Pro Gly Thr Lys Leu Glu Ile Lys Arg
                225                 230
```

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 109 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

```
Glu Thr Thr Leu Thr Gln Ser Pro Gly Thr Leu Ser Leu Ser Pro Gly
1               5                   10                  15

Glu Arg Ala Thr Leu Ser Cys Arg Ala Ser Gln Asn Ile Gly Ser Ser
            20                  25                  30

Ser Leu Ala Trp Tyr Gln Gln Lys Pro Gly Gln Ala Pro Arg Leu Leu
        35                  40                  45

Ile Tyr Gly Ala Ser Thr Arg Ala Thr Gly Ile Pro Ala Arg Phe Ser
    50                  55                  60

Gly Ser Gly Ser Gly Thr Gln Phe Thr Leu Thr Ile Ser Ser Leu Gln
65                  70                  75                  80

Ser Glu Asp Phe Ala Val Tyr Tyr Cys Gln Gln Tyr Asn Phe Trp Pro
                85                  90                  95

Phe Thr Phe Gly Pro Gly Thr Lys Leu Glu Ile Lys Arg
            100                 105
```

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 72 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

TTTTCTGCAG TAATAGGCGG CCGCAGGGGG AGGAGGGTCC ATCGAAGGTC GCGAAGCAGA 60

GACTGTTGAA AG 72

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 75 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

TTTTGAATTC TTATTAACCA CCGAACTGCG GGTGACGCCA AGCGCTTGCG GCCGTTAAGA 60

CTCCTTATTA CGCAG 75

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

AAAAGCGGCC GCACTGGCCT GAGAGANNNN NN 32

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

TCGACCCACG CGTCCG 16

(2) INFORMATION FOR SEQ ID NO: 23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

GGGTGCCGAG GC 12

(2) INFORMATION FOR SEQ ID NO: 24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 72 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

AAAAGAATTC TGAATGTGTT ATTTTAGCTC AGGTCACTCT GACCTGATTA TCAAGACCCC 60

ACTTCACCCC CT 72

(2) INFORMATION FOR SEQ ID NO: 25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 46 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

AAAAGCGGCC GCAGGGGGAG GAGGGTCCAT GGAATCTGCC AAGGAG 46

(2) INFORMATION FOR SEQ ID NO: 26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

GTAAAACGAC GGCCAGT 17

(2) INFORMATION FOR SEQ ID NO: 27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

GGATAACAAT TTCACACAGG 20

(2) INFORMATION FOR SEQ ID NO: 28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

AAAGCGGCCG CACTGGCCTG AGAGA 25

(2) INFORMATION FOR SEQ ID NO: 29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 41 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

AAAAGGCCCA GCCGGCCATG GCCCAGCCCA CCACGCGTCC G 41

(2) INFORMATION FOR SEQ ID NO: 30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

AAAAGGCCCA GCCGGCCATG GCCCAGTCCC ACCACGCGTC CG 42

(2) INFORMATION FOR SEQ ID NO: 31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 43 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

AAAAGGCCCA GCCGGCCATG GCCCAGTACC CACCACGCGT CCG 43

(2) INFORMATION FOR SEQ ID NO: 32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 72 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

AAAAGAATTC GAGGTTACAT TAACTTTGTT CCGGTCAGAC TGACCCAAGT CGACCTGAAT 60

GTGTTATTTT AG 72

(2) INFORMATION FOR SEQ ID NO: 33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

CCTTGGCAGA TTCCATCTCG GCCATTGCCG GC 32

(2) INFORMATION FOR SEQ ID NO: 34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

CCGGCAATGG CCGAGATGGA ATCTGCCAAG G 31

(2) INFORMATION FOR SEQ ID NO: 35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 86 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

TTTTGTCGAC TCAATCAGTT ATGCGGCCGC CAGCTGCAGG AGGGCCGGCT GGGCCGACCC 60

TCCTCCCCCA GACCCCACTT CACCCC 86

(2) INFORMATION FOR SEQ ID NO: 36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

TGTTGAAACA CAAGCGCCAG 20

(2) INFORMATION FOR SEQ ID NO: 37:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

TGGAAGAGGC ACGTTCTTTT CTTT 24

(2) INFORMATION FOR SEQ ID NO: 38:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

CTCCTTCTTA CTCTTGCTGG CGGT 24

(2) INFORMATION FOR SEQ ID NO: 39:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

AGACTCTCCC CTGTTGAAGC TCTT 24

(2) INFORMATION FOR SEQ ID NO: 40:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

TGAAGATTCT GTAGGGGCCA CTGTCTT 27

(2) INFORMATION FOR SEQ ID NO: 41:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 41:

TGAACCGCCT CCACCTGAGG AGACGGTGAC CAGGGTGCC 39

(2) INFORMATION FOR SEQ ID NO: 42:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 42:

TGAACCGCCT CCACCTGAAG AGACGGTGAC CATTGTCCC 39

(2) INFORMATION FOR SEQ ID NO: 43:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 43:

TGAACCGCCT CCACCTGAGG AGACGGTGAC CAGGGTTCC 39

(2) INFORMATION FOR SEQ ID NO: 44:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 44:

TGAACCGCCT CCACCTGAGG AGACGGTGAC CGTGGTCCC 39

(2) INFORMATION FOR SEQ ID NO: 45:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 47 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 45:

TTTTTGGCCC AGCCGGCCAT GGCCCAGGTG CAGCTGGTGC AGTCTGG 47

(2) INFORMATION FOR SEQ ID NO: 46:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 47 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 46:

TTTTTGGCCC AGCCGGCCAT GGCCCAGGTC AACTTAAGGG AGTCTGG 47

(2) INFORMATION FOR SEQ ID NO: 47:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 47 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 47:

TTTTTGGCCC AGCCGGCCAT GGCCGAGGTG CAGCTGGTGG AGTCTGG 47

(2) INFORMATION FOR SEQ ID NO: 48:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 47 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 48:

TTTTTGGCCC AGCCGGCCAT GGCCCAGGTG CAGCTGCAGG AGTCGGG 47

(2) INFORMATION FOR SEQ ID NO: 49:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 47 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 49:

TTTTTGGCCC AGCCGGCCAT GGCCGAGGTG CAGCTGTTGC AGTCTGC 47

(2) INFORMATION FOR SEQ ID NO: 50:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 47 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 50:

TTTTTGGCCC AGCCGGCCAT GGCCCAGGTA CAGCTGCAGC AGTCAGG 47

(2) INFORMATION FOR SEQ ID NO: 51:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 51:

TTATTCGCGG CCGCCTAAAC AGAGGCAGTT CCAGATTTC 39

(2) INFORMATION FOR SEQ ID NO: 52:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 52:

GTCACTTGCG GCCGCCTACA GTGTGGCCTT GTTGGCTTG 39

(2) INFORMATION FOR SEQ ID NO: 53:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 41 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 53:

TCTGGCGGTG GCGGATCGGA CATCCAGATG ACCCAGTCTC C 41

(2) INFORMATION FOR SEQ ID NO: 54:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 41 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 54:

TCTGGCGGTG GCGGATCGGA TGTTGTGATG ACTCAGTCTC C 41

(2) INFORMATION FOR SEQ ID NO: 55:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 41 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 55:

TCTGGCGGTG GCGGATCGGA AATTGTGTTG ACGCAGTCTC C 41

(2) INFORMATION FOR SEQ ID NO: 56:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 41 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 56:

TCTGGCGGTG GCGGATCGGA CATCGTGATG ACCCAGTCTC C 41

(2) INFORMATION FOR SEQ ID NO: 57:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 41 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 57:

TCTGGCGGTG GCGGATCGGA AACGACACTC ACGCAGTCTC C 41

(2) INFORMATION FOR SEQ ID NO: 58:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 41 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 58:

TCTGGCGGTG GCGGATCGGA AATTGTGCTG ACTCAGTCTC C 41

(2) INFORMATION FOR SEQ ID NO: 59:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 59:

TTCTCGTGCG GCCGCCTAAC GTTTGATTTC CACCTTGGTC CC 42

(2) INFORMATION FOR SEQ ID NO: 60:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 60:

TTCTCGTGCG GCCGCCTAAC GTTTGATCTC CAGCTTGGTC CC 42

(2) INFORMATION FOR SEQ ID NO: 61:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 61:

TTCTCGTGCG GCCGCCTAAC GTTTGATATC CACTTTGGTC CC 42

(2) INFORMATION FOR SEQ ID NO: 62:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 62:

```
TTCTCGTGCG GCCGCCTAAC GTTTGATCTC CACCTTGGTC CC                         42
```

(2) INFORMATION FOR SEQ ID NO: 63:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 63:

```
TTCTCGTGCG GCCGCCTAAC GTTTAATCTC CAGTCGTGTC CC                         42
```

(2) INFORMATION FOR SEQ ID NO: 64:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 41 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 64:

TCTGGCGGTG GCGGATCGCA GTCTGTGTTG ACGCAGCCGC C
41

(2) INFORMATION FOR SEQ ID NO: 65:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 41 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 65:

TCTGGCGGTG GCGGATCGCA GTCTGCCCTG ACTCAGCCTG C
41

(2) INFORMATION FOR SEQ ID NO: 66:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 41 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthesis DNA"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 66:

TCTGGCGGTG GCGGATCGTC CTATGTGCTG ACTCAGCCAC C
41

(2) INFORMATION FOR SEQ ID NO: 67:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 41 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthesis DNA"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 67:

TCTGGCGGTG GCGGATCGTC TTCTGAGCTG ACTCAGGACC C 41

(2) INFORMATION FOR SEQ ID NO: 68:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 41 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthesis DNA"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 68:

TCTGGCGGTG GCGGATCGCA CGTTATACTG ACTCAACCGC C 41

(2) INFORMATION FOR SEQ ID NO: 69:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 41 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthesis DNA"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 69:

TCTGGCGGTG GCGGATCGCA GGCTGTGCTC ACTCAGCCGT C 41

(2) INFORMATION FOR SEQ ID NO: 70:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 41 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthesis DNA"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 70:

```
TCTGGCGGTG GCGGATCGAA TTTTATGCTG ACTCAGCCCC A                          41
```

(2) INFORMATION FOR SEQ ID NO: 71:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 71:

```
TTCTCGTGCG GCCGCCTAAC CTAGGACGGT GACCTTGGTC CC                         42
```

(2) INFORMATION FOR SEQ ID NO: 72:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 72:

```
TTCTCGTGCG GCCGCCTAAC CTAGGACGGT CAGCTTGGTC CC                         42
```

(2) INFORMATION FOR SEQ ID NO: 73:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 73:

TTCTCGTGCG GCCGCCTAAC CTAAAACGGT GAGCTGGGTC CC 42

(2) INFORMATION FOR SEQ ID NO: 74:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 74:

CGATCCGCCA CCGCCAGA 18

(2) INFORMATION FOR SEQ ID NO: 75:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 75:

GTCTCCTCAG GTGGAGGC 18

(2) INFORMATION FOR SEQ ID NO: 76:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 54 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 76:

CGATCCGCCA CCGCCAGAGC CACCTCCGCC TGAACCGCCT CCACCTGAGG AGAC 54

Analytical Biochemistry **294**, 108–117 (2001)
doi:10.1006/abio.2001.5147, available online at http://www.idealibrary.com on IDEAL®

External Surface Display of Proteins Linked to DNA-Binding Domains

Duncan P. McGregor and Simon P. Robins
Skeletal Metabolism Research Unit, Rowett Research Institute, Greenburn Road, Bucksburn, Aberdeen, Scotland, AB21 9SB

Received September 26, 2000; published online June 15, 2001

A novel system (DBDX) was developed which allows the external surface display on filamentous bacteriophage of proteins fused to either the N- or the C-terminus of a DNA-binding protein. In conjunction with helper phage infection, expression of proteins fused to the estrogen receptor DNA-binding domain (DBD) in a phagemid vector containing the DNA sequence recognized by the DBD resulted in the production of phage particles which display the fusion protein through the phage pVIII coat on the external surface of the particle. The viability of the technique was established with several model systems: particles displaying the C-terminal domain of N-cadherin or the biotinylation domain of propionyl coenzyme A carboxylase fused to the C-terminus of the DBD were found to be bound specifically by antibody or streptavidin, respectively. Human κ constant region cDNA was selected from a N-terminal DBD fusion lymphocyte cDNA library after two rounds of selection with anti-κ antibody. This display system may complement currently available bacterial selection techniques. © 2001 Academic Press

Key Words: **protein engineering; DNA-binding protein; display systems; library screening.**

All biological display systems must link phenotype with genotype in order to identify DNA encoding a peptide or protein with a desired property. Several different methods have been developed to achieve this. For example, in phage display systems (1), libraries of proteins are fused to various filamentous bacteriophage coat proteins, with the DNA encoding the fusion protein packaged within the viral particle (Fig. 1). Libraries of fusion proteins incorporated into phage can be selected against targets of interest and, following amplification of the bound phage, the selection is repeated, resulting in the enrichment of binding proteins. Peptides (2–4), antibody fragments (5–7), and proteins (8, 9) have successfully been displayed as N-terminal pIII fusions and, to a much lesser extent, as N-terminal pVIII, pVII, and pIX fusions (10, 11) or as C-terminal pVI fusions (12).

An alternative method is based on the DNA-binding ability of the lac repressor (13, 14). Libraries of random peptides are fused to the C-terminal end of the *lacI* repressor protein through expression from a plasmid vector carrying the fusion gene. Linkage of the *LacI*–peptide fusion to its encoding DNA occurs via the *lacO* sequences on the plasmid, forming a stable plasmid–*LacI*–peptide complex (Figure 1). These complexes are released from their host bacteria by cell lysis, and peptides of interest are isolated by affinity purification on an immobilized target. The plasmids thus isolated can then be reintroduced into *Escherichia coli* by electroporation to amplify the selected population for additional rounds of screening.

Bacteriophage pIII fusion display is the most commonly used and robust of these systems, but it does have some limitations relating to the orientation of the displayed protein (N-terminal fusion only) and the size of the displayed protein.

We set out to develop a versatile display system (DBDX) that did not require fusion to a coat protein while still displaying a protein to the external environment. It seemed possible that if a DNA-binding domain could remain bound to single stranded phagemid DNA during packaging, a protein fused to the DNA-binding domain (DBD)[1] could protrude through the protein

---

[1] Abbreviations used: DBD, DNA-binding domain; HRE, DNA-binding domain; k.r.u., kanamycin resistance units; IPTG, isopropyl β-D-thiogalactoside; PEG, polyethylene glycol; PBS, phosphate-buffered saline; PCCα propionyl coenzyme A α; ELISA, enzyme-linked immunosorbent assay; IgG, immunoglobulin G; BSA, bovine serum albumin; BPBS, BSA in PBS; MPBS, 2% milk powder (Marvel)/PBS; HRP, horseradish peroxidase; UTR, untranslated region.

0003-2697/01 $35.00
Copyright © 2001 by Academic Press
All rights of reproduction in any form reserved.

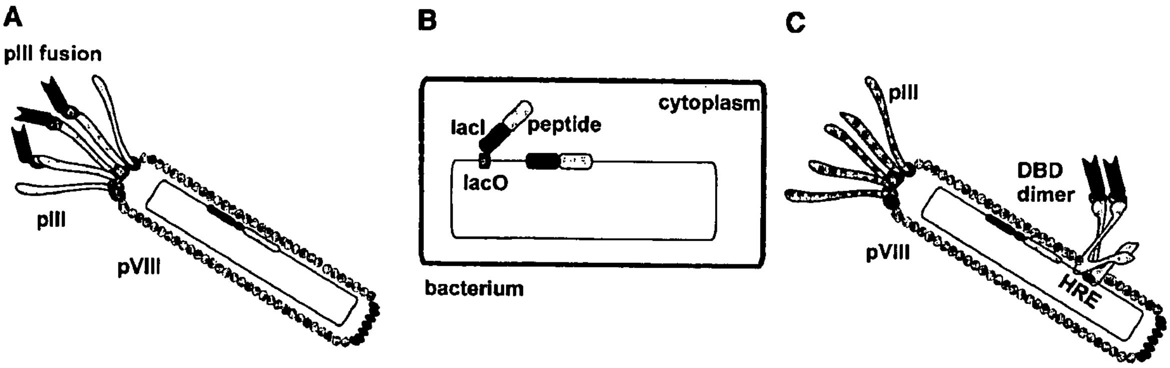

FIG. 1. Alternative bacterial display systems are shown schematically, highlighting the differences between the phage display (pIII fusion) (A), the lacI C-terminal "peptides on plasmids" (B), and the external surface DBDX (C) display systems. Phage pIII and pVIII proteins are labeled in (A) and (C). pIII, lacI, and DBD fusions are shown as protein bound to DNA or displayed on the viral particle and also shown encoded on the vector DNA. For the DBDX figure, dimeric DBD-fusion protein is shown "bound" to the vector DNA and protruding through the protein pVIII coat.

coat allowing selection of binding proteins from libraries by standard panning methodologies (Fig. 1).

The DBD of the estrogen receptor was chosen for the development of this system for a number of reasons. Although the estrogen receptor is a large multifunctional polypeptide of 595 amino acids which functions in the cytoplasm and nucleus of eukaryotic cells (15), a minimal high-affinity DBD has been defined between amino acids 176 and 282 and expressed in *E. coli* (16). In addition, mutated progesterone receptor DBD has been shown to bind to the consensus estrogen hormone response element (HRE) sequence in bacterial cytoplasm (17). The DNA-binding function of the estrogen receptor DBD is not inhibited by the presence of non-DNA-binding domains at either the N- or the C-terminal ends of this domain in the full-length protein. We therefore wanted to investigate the potential for both N- and C-terminal display using the same DBD.

It was anticipated that the estrogen DBD would bind as a dimer to the specific, double-stranded DNA estrogen receptor target HRE nucleotide sequence with a similar affinity (0.5 nM) to that of the parent molecule (18, 19). DBD dimerization on the surface of the particle would result in a maximum of two peptides displayed per particle. Additionally, the estrogen receptor DBD is capable of binding to its 38-bp target HRE sequence with high affinity and specificity under the salt and pH conditions normally required for selection of binding peptides (19). The core 13-bp interrupted palindromic repeat of the HRE needs to be present to allow high-affinity DBD binding, with the remainder of the HRE sequence stabilizing binding (18). Moreover, binding affinity is apparently increased 60-fold for the single-stranded coding, or "plus," strand of the HRE nucleotide sequence (to 10^{-11} M) over the double-stranded form (20). This preference for one strand of the HRE sequence seemed ideally suited to a phagemid/helper phage packaging system.

N- and C-terminal display vectors were constructed and used to demonstrate the principle with individual proteins. A human lymphocyte cDNA library was then constructed and screened to demonstrate that selection, as well as display, was also possible using this system.

MATERIALS AND METHODS

Oligonucleotide sequences used for vector construction are shown in Table 1.

Display vector construction. The N-terminal display vector pDM12 (Fig. 2) was constructed by inserting a *Not*I–*Eco*RI-digested estrogen receptor DNA-binding domain, modified by appropriate PCR primers, into a phagemid vector pDM6 which is based on the pUC119-derived phage display vector, pHEN1 (21). The estrogen receptor DBD was isolated from double-stranded cDNA prepared from human bone marrow mRNA (Clontech, Palo Alto, CA) using a Superscript plasmid cDNA synthesis kit (Life Technologies Ltd., Paisley, Scotland, UK). The estrogen receptor was PCR amplified from 5 μl (150–250 ng) of bone marrow cDNA using 25 pmol of each of the primers pDM12FOR and pDM12BAK in two 50-μl reactions containing 0.1 mM dNTPs, 2.5 units *Taq* DNA polymerase, and 1× PCR reaction buffer (10 mM Tris–HCl, pH 9.0, 5 mM KCl, 0.01% Triton X-100, 1.5 mM $MgCl_2$) (Promega Ltd., Southampton, UK). The pDM12BAK primer anneals to the 5′-end of the DNA-binding domain of the estrogen receptor and incorporates the $(Gly)_4$Ser linker and the *Not*I restriction site. The pDM12FOR primer anneals to the 3′-end of the DNA-binding domain of the estrogen receptor and incorporates two stop codons, the 38-bp consensus estrogen receptor HRE (18), and an *Eco*RI restriction site.

PCRs were carried out on a Techne PHC-3 thermal cycler for 30 cycles of 94°C, 1 min; 65°C, 1 min; 72°C, 1

TABLE 1

Oligonucleotide Primers Used for Sequencing, Vector Construction, Insert Transfer, and Library Construction

Name	Sequence
M13FOR	5′-GTAAAACGACGGCCAGT
M13REV	5′-GGATAACAATTTCACACAGG
ORSEQBAK	5′-TGTTGAAACACAAGCGCCAG
ORSEQFOR	5′-GCACACTGCACAGTAGCGAG
CDNAFOR	5′-AAAAGCGGCCGCACTGGCCTGAGAGA $(N)_6$
CDNAPCRFOR	5′-AAAGCGGCCGCACTGGCCTGAGAGA
CDNAPCRBAK1	5′-AAAAGGCCCAGCCGGCCATGGCCCAGCCCACCACGCGTCCG
CDNAPCRBAK2	5′-AAAAGGCCCAGCCGGCCATGGCCCAGTCCCACCACGCGTCCG
CDNAPCRBAK3	5′-AAAAGGCCCAGCCGGCCATGGCCCAGTACCCACCACGCGTCCG
pDM12FOR	5′-AAAAGAATTCTGAATGTGTTATTTTAGCTCAGGTCACTCTGACCTGATTATCAAGACCCCACTTCACCCCCT
pDM12BAK	5′-AAAAGCGGCCGCAGGGGGAGGAGGGTCCATGGAATCTGCCAAGGAG
pDM14FOR	5′-AAAAGAATTCGAGGTTACATTAACTTTGTTCCGGTCAGACTGACCCAAGTCGACCTGAATGTGTTATTTTAG
pDM15FOR	5′-AAAAGAATTCTTATTAGTGGTGGTGGTGGTGGTGAGACCCCACTTCACCCCTG
pDM16FOR	5′-TTTTGTCGACTCAATCAGTTATGCGGCCGCCAGCTGCAGGAGGGCCGGCTGGGCCGACCCTCCTCCCCCAGACCCCACTTCACCCC
PELBBAK	5′-CCGGCAATGGCCGAGATGGAATCTGCCAAGG
PELBFOR	5′-CCTTGGCAGATTCCATCTCGGCCATTGCCGGC

min. Vector- and gel-purified PCR fragment DNAs were *Not*I–*Eco*RI cleaved and ligated together and then electroporated into TG-1 *E. coli* cells (genotype, K12, (Δlac-pro), supE, thi, hsD5/F′traD36, $proA^+B^+$, $LacI^q$, LacZΔ15) and plated onto 2×TY agar plates supplemented with 1% glucose and 100 μg/ml ampicillin.

The pDM14 and pDM15 vectors were constructed by PCR amplification with M13REV and pDM14FOR and pDM15FOR, respectively, using 20 ng pDM12 vector DNA as template, using the conditions described above. Gel-purified PCR products were then *Not*I–*Eco*RI cleaved and ligated into *Not*I–*Eco*RI-cleaved pDM12 vector DNA to produce the two new vectors. The pDM14 vector contains two HRE sites separated by a *Sal*I site and was used in the construction of the pDM16 vector in this study. The pDM15 vector lacks any HRE site and was used to demonstrate the requirement of such a site for external surface display. Key regions of both vectors are shown in Fig. 3.

For the pDM16 C-terminal display vector, the pelB leader DNA sequence was fused directly to the estrogen receptor DBD sequence, removing the multiple cloning sites and the Gly_4Ser linker DNA sequence found in pDM12, which were appended to the C-terminal end of the DBD sequence upstream of the HRE DNA sequence. The vector was constructed by joining the products of two separate PCR reactions by overlapping PCR (22). To create this vector, PCR reactions were carried out on a Techne Progene thermal cycler for 30 cycles of 94°C, 1 min; 60°C, 1 min; 72°C, 1 min. In the first reaction, the 5′-untranslated region and pelB leader DNA sequence was amplified from 100 ng of pDM12 vector DNA using 50 pmol of each of the oligonucleotides, M13REV and pelBFOR, in a 100-μl reaction containing 0.1 mM dNTPs, 2.5 units *Taq*plus DNA polymerase, and 1× high salt PCR reaction buffer (20 mM Tris–HCl, pH 9.2, 60 mM KCl, 2 mM $MgCl_2$) (Stratagene Ltd., Cambridge, U.K.). In the second reaction, the 3′-end of the pelB leader sequence and the estrogen receptor DBD were amplified from 100 ng of pDM12 vector DNA using 50 pmol of each of the oligonucleotides, pelBBAK and pDM16FOR, in a 100-μl reaction containing 0.1 mM dNTPs, 2.5 units *Taq*plus DNA polymerase, and 1× high salt PCR buffer. Following gel purification both products were mixed together and a final round of PCR amplification with *Taq* DNA polymerase (Promega Ltd., Southampton, UK) was performed with M13REV and pDM16FOR. The resulting DNA fragment was *Hin*dIII–*Sal*I restriction enzyme digested and cloned into *Hin*dIII–*Sal*I-digested pDM14 vector to create the pDM16 vector.

All new vector insert sequences were confirmed by sequencing with a Prism dyedeoxy cycle sequencing kit (Perkin–Elmer, Warrington, Lancashire, UK) using M13FOR and M13REV oligonucleotides.

N-terminal display of leptin. Amino acids V22–C167 of placental human leptin cDNA (23) had previously been cloned as a *Sfi*I–*Not*I fragment into a pIII phagemid display vector (unpublished data). *Sfi*I–*Not*I-digested leptin cDNA was cloned into the pDM12 and pDM15 vectors, electroporated into TG-1 *E. coli* bacteria, and plated onto 2×TY agar plates supplemented with 1% glucose and 100 μg/ml ampicillin. Leptin-producing pDM12 clones were detected by phage ELISA of individual culture supernatants with protein G-purified rabbit polyclonal anti-leptin IgG kindly provide by Dr. Nigel Hoggard (Rowett Research Institute) (not shown). Leptin cDNA containing pDM15 clones were identified by PCR amplification with M13REV and M13FOR, using 1 μl of an overnight culture as template. ELISA-positive pDM12 clones were sequenced with the M13REV and ORSEQFOR primers (Table 1). Leptin cDNA-containing pDM15 clones were sequenced with the M13REV and M13FOR primers to confirm insertion of leptin cDNA in frame with the leader sequence and the estrogen receptor DBD. Five milliliters of 2×TY culture broth supplemented with 1% glucose and 100 μg/ml ampicillin was inoculated with either pDM12–leptin or pDM15–leptin–DBD fusion clone TG-1 stocks and grown for 5 h at 37°C until optical density at 600 nm ($OD_{600\ nm}$) reached 0.4. The cultures were infected with 10^9 kanamycin resistance units (k.r.u.) M13K07 helper phage for 1 h at 37°C. A 1-ml aliquot of each of the infected cultures was transferred to fresh tubes and adjusted to 5 ml with 2×TY broth supplemented with 25 μg/ml kanamycin, 100 μg/ml ampicillin, and 20 μM IPTG. Amplification was carried out overnight at 37°C, and phages were then PEG/NaCl precipitated from 2 ml of culture and resuspended in 500 μl of PBS. The volume was adjusted to 1 ml with 4% Marvel/PBS and 100 μl/well tested in ELISA against anti-human leptin antibody as described below.

C-terminal display of N-cadherin and propionyl coenzyme A carboxylase α (PCCα). Fragments of both cDNAs had previously been cloned from a pVI fusion C-terminal display oligo(dT)–*Not*I-primed SaOS-2 osteosarcoma-cell-line cDNA library (unpublished data). These clones contain an adaptor sequence at the 5′-end of the cDNA fragment, which could be modified by PCR to allow in-frame cloning of these cDNAs as *Sfi*I–*Not*I fragments into pDM16. A 1.4-kb human N-cadherin cDNA fragment encoding the C-terminal 99 amino acids, the 3′-untranslated region, and the poly(A) tail and a 0.8-kb human cDNA fragment of PCCα encoding the C-terminal 95 amino acids, the 3′-untranslated region, and the poly(A) tail were amplified from approximately 20 ng insert containing vector with oligonucleotides M13FOR and CDNAPCRBAK1. Following gel purification and *Sfi*I–*Not*I digestion, the PCR products were cloned into pDM16 and sequenced with M13FOR and ORSEQBAK primers. Fiftyfold concentrated PBS stocks of C-terminal N-cadherin and PCCα phage particles were prepared from 20 ml overnight helper-phage-infected cultures grown at 30°C as described for the cDNA library stock preparation.

Preparation of individual clones for ELISA. Individual colonies were picked into 100 μl 2×TY (24) supplemented with 100 μg/ml ampicillin and 1% (w/v) glucose in 96-well tissue culture plates and incubated at 37°C with shaking at 200 rpm for 4–6 h. For each culture, 25 μl was transferred to a fresh 96-well plate containing 25 μl/well of the same medium plus 10^7–10^8 k.r.u. of the M13K07 kanamycin-resistant helper phage and incubated at 37°C for 1 h. Expression medium (160 μl/well 2×TY/100 μg/ml ampicillin/25 μg/ml kanamycin/20 μM IPTG) was added to each well and amplification continued overnight at 30°C (37°C for round 2 cDNA library clones). The cultures were centrifuged (2000*g*, 10 min) at 4°C and 50μl of culture supernatant was used for ELISA.

ELISA screening. Immulon 4 ELISA plates (Dynatech Laboratories) were coated at 37°C for 1 h with 100 μl/well 1:1500 diluted rabbit polyclonal anti-leptin IgG, 1:250 diluted anti-cadherin monoclonal antibody, 2.5 μg/ml streptavidin (Sigma Chemical Co., Poole, Dorset, UK), or 2 μg/ml anti-human κ antibody in PBS. The plates were blocked for 1 h at room temperature with either 2% (w/v) BSA in PBS (BPBS) for coating with anti-κ antibody or with 2% milk powder (Marvel)/PBS (MPBS), for coatings with anti-leptin antibody, anti-cadherin antibody, or streptavidin. Culture supernatant or concentrated phage were incubated for 1 h at room temperature in PBS containing 2% BSA and 0.1% (v/v) Tween 20 (κ chain clones) or 2% MPBS. Bound particles were detected with 1:5000 diluted anti-M13–HRP conjugate (Amersham-Pharmacia Ltd., Little Chalfont, Buckinghamshire, UK) in the same buffer. After each incubation, plates were washed three times with 200 μl/well PBS containing 0.1% Tween 20, and three times more with PBS. Plates were developed for 5–10 min at room temperature with 3,3′,5,5′-tetramethylbenzidine substrate buffer (Sigma Chemical Co.) and the optical densities were read at 450 nm.

Lymphocyte cDNA library construction. mRNA was isolated from approximately 10^9 human peripheral blood lymphocytes using a poly(ATract) 1000 mRNA isolation kit (Promega, Southampton, UK). Randomly primed double-stranded cDNA was synthesized with a Superscript plasmid cDNA synthesis kit (Life Technologies Ltd.) using 5 μg of lymphocyte mRNA and 50 pmol CDNAFOR primer. cDNA was PCR amplified using 25 pmol per reaction of oligonucleotides CDNAPCRFOR and an equimolar mixture of CDNAPCRBAK1-3, which anneal to the 5′-end cDNA adaptor sequence and incorporate a *Sfi*I restriction site. Ten 50-μl PCR reactions were carried out using 2 μl of cDNA (50 ng) per reaction, as described above, using *Taq* DNA polymerase, for 25 cycles of 94°C, 1 min; 60°C, 1 min; 72°C, 2 min. Reaction mixtures were pooled, phenol/chloroform extracted, and ethanol precipitated. Amplified cDNA and 5 μg of pDM12 vector DNA were *Sfi*I–*Not*I digested and small DNA fragments were removed by size selection on Chromaspin 1000 spin columns (Clontech, Palo Alto, CA). Ligation was carried out overnight at 16°C. The ligated DNA was electroporated into TG-1 *E. coli*. Cells were plated onto 90-mm 2×TY agar plates supplemented with 1%

glucose and 100 μg/ml ampicillin and allowed to grow at 30°C overnight. Dilutions (10^{-4}, 10^{-5}, and 10^{-6}) of the electroporated bacteria were also plated to assess library size. Forty milliliters of 2×TY broth supplemented with 20% glycerol, 1% glucose, and 100 μg/ml ampicillin was poured onto the bacterial lawns and the bacteria were then scraped off the plates. A 5-ml aliquot was added to a 20-ml 2×TY culture broth supplemented with 1% glucose and 100 μg/ml ampicillin and then infected with 10^{11} k.r.u. M13K07 helper phage at 37°C for 1 h. Infected bacteria were transferred to 200 ml 2×TY broth supplemented with 25 μg/ml kanamycin, 100 μg/ml ampicillin, and 20 μM IPTG and incubated overnight at 37°C. Bacteria were removed by centrifugation at 2000*g* for 20 min in 50-ml Falcon tubes. Forty milliliters of 2.5 M NaCl/20% (w/v) PEG 6000 was added to 200 ml of culture supernatant, mixed vigorously, and incubated on ice for 1 h to precipitate phage. Particles were pelleted at 11,000 rpm for 30 min in 250-ml centrifuge tubes at 4°C in a Sorvall RC5B centrifuge and resuspended in PBS buffer (2 ml), after removing all traces of PEG/NaCl with a pipette. Bacterial debris was then removed by centrifugation followed by filtration of the supernatant through a 0.45-μm polysulfone syringe filter (Gelman Sciences, Ann Arbor, MI) and stored at −20°C.

cDNA library screening. For the first round of library selection, a 70 × 11-mm Nunc Maxisorp immunotube (Life Technologies Ltd.) was coated with 2.5 ml of 10 μg/ml anti-human κ antibody (Harlan-Seralab, Loughborough, Leicestershire, UK) in PBS for 2 h at 37°C. The tube was rinsed three times with PBS and blocked with BPBS (3 ml) for 2 h at 37°C and washed as described above. In the first round of selection, primary pDM12–lymphocyte cDNA phage stock was incubated in PBS containing 2% BSA and 0.05% Tween 20 for 2 h at room temperature. The tube was washed 10 times with PBS containing 0.1% Tween 20, followed by 10 washes with PBS only. Bound particles were eluted in freshly prepared 0.1 M triethylamine (1 ml) for 10 min at room temperature on a blood mixer. Eluted particles were transferred to 0.5 ml 1 M Tris, pH 7.4, vortex mixed briefly, and transferred to ice. Neutralized eluates were added to 10-ml log-phase TG1 bacteria ($OD_{600\ nm}$ 0.3–0.5) and incubated at 37°C for 1 h. Dilutions (10^{-3}, 10^{-4}, and 10^{-5}) of the infected culture were prepared to estimate the number of particles recovered. The remainder of the culture was spun at 2000*g* for 10 min, and the pellet was resuspended in 300 μl 2×TY medium by vortex mixing. The infected bacteria were plated onto 2×TY agar plates supplemented with 1% glucose and 100 μg/ml ampicillin. The round 2 selection was carried out with 250 μl of 100-fold concentrated round 1 amplified stock prepared from a 100-ml overnight helper-phage-infected culture as described above for the primary library amplification, with the number of PBS–Tween and PBS washes increased to 12. Individual round 2 clones were then tested in ELISA against anti-human κ antibody, and inserts were identified by sequencing with M13REV primer.

Effect of IPTG concentration and growth temperature. Five milliliters of 2×TY culture broth supplemented with 1% glucose and 100 μg/ml ampicillin was inoculated with a κ constant region–DBD fusion clone TG-1 stock and grown for 5 h at 37°C until $OD_{600\ nm}$ reached 0.4. The culture was infected with 10^9 k.r.u. M13K07 helper phage for 1 h at 37°C. Aliquots (0.5 ml) of the infected culture were transferred to six separate tubes and adjusted to 5 ml with 2×TY broth supplemented with 25 μg/ml kanamycin and 100 μg/ml ampicillin. The IPTG concentration was adjusted to 20, 50, or 100μM in duplicate samples. Amplification was carried out overnight at 30 or 37°C for each IPTG concentration. Phages were then PEG/NaCl precipitated from 2 ml of culture and resuspended in 800 μl of PBS, adjusted to 1600 μl with 4% BSA/0.2% Tween 20/PBS, and tested in ELISA against anti-human κ antibody.

RESULTS

The aim of this study was to develop a system of display that complements existing phage display and "peptides on plasmids" methods. The phagemid vectors constructed here (Figs. 2 and 3) allow both N- and C-terminal fusion of inserts to the estrogen receptor DNA-binding domain as opposed to N-terminal display for pIII and pVIII phage display and C-terminal display for pVI and lacI fusions. Estrogen receptor DBD binding to DNA should not be disrupted significantly by the presence of non-DBD sequence at the N- or C-terminus of the DBD. The HRE sequence present in the two display vectors used in this study, pDM12 and pDM16, differ slightly as the pDM16 HRE was derived from pDM14 (Fig. 3). The pDM15 vector (Fig. 3) was constructed to allow expression and purification of estrogen receptor fusion proteins and to test for the absolute requirement of an HRE site for external surface display. The vectors are designed for fusion protein expression in conjunction with helper phage infection to allow optimal binding to the single-stranded "plus" HRE sequence during packaging of phagemid DNA into mature phage particles.

The first stage of the present study was to show that individual proteins or fragments of proteins could be displayed through the bacteriophage pVIII coat and that binding to target could be detected in concentrated samples and crude culture supernatant. Detection of phage binding in culture supernatant is essential for screening large numbers of clones after library selec-

```
M13REV ────→
AACAGCTATGACCATG        HindIII
                                                                          pelB
                                                                 MET LYS TYR LEU LEU PRO THR ALA ALA ALA GLY LEU
AACAGCTATGACCATGATTACGCCAAGCTTGCATGCAAATTCTATTTCAAGGAGACAGTCATAA ATG AAA TAC CTA TTG CCT ACG GCA GCC GCT GGA TTG
TTGTCGATACTGGTACTAATGCGGTTCGAACGTACGTTTAAGATAAAGTTCCTCTGTCAGTATT TAC TTT ATG GAT AAC GGA TGC CGT CGG CGA CCT AAC

                SfiI                                PstI            NotI
LEU LEU LEU ALA ALA GLN PRO ALA MET ALA                     *** *** ALA ALA ALA|GLY GLY GLY GLY SER|MET GLU SER
TTA TTA CTC GCG GCC CAG CCG GCC ATG GCC GAG GTG CAA CTG CAG TAA TAG GCG GCC GCA|GGG GGA GGA GGG TCC|ATG GAA TCT
AAT AAT GAG CGC CGG GTC GGC CGG TAC CGG GTC CAC GTC GAC GTC ATT ATC CGC CGG CGT|CCC CCT CCT CCC AGG|TAC CTT AGA
      ←──── ORSEQFOR
             GA GCG ATG ACA CGT CAC ACG                Oestrogen receptor DBD aa 176-282
ALA LYS GLU THR ARG TYR CYS ALA VAL CYS ASN ASP TYR ALA SER GLY TYR HIS TYR GLY VAL TRP SER CYS GLU GLY CYS LYS
GCC AAG GAG ACT CGC TAC TGT GCA GTG TGC AAT GAC TAT GCT TCA GGC TAC CAT TAT GGA GTC TGG TCC TGT GAG GGC TGC AAG
CGG TTC CTC TGA GCG ATG ACA CGT CAC ACG TTA CTG ATA CGA AGT CCG ATG GTA ATA CCT CAG ACC AGG ACA CTC CCG ACG TTC

ALA PHE PHE LYS ARG SER ILE GLN GLY HIS ASN ASP TYR MET CYS PRO ALA THR ASN GLN CYS THR ILE ASP LYS ASN ARG ARG
GCC TTC TTC AAG AGA AGT ATT CAA GGA CAT AAC GAC TAT ATG TGT CCA GCC ACC AAC CAG TGC ACC ATT GAT AAA AAC AGG AGG
CGG AAG AAG TTC TCT TCA TAA GTT CCT GTA TTG CTG ATA TAC ACA GGT CGG TGG TTG GTC ACG TGG TAA CTA TTT TTG TCC TCC

LYS SER CYS GLN ALA CYS ARG LEU ARG LYS CYS TYR GLU VAL GLY MET MET LYS GLY GLY ILE ARG LYS ASP ARG ARG GLY GLY
AAG AGC TGC CAG GCC TGC CGG CTC CGT AAA TGC TAC GAA GTG GGA ATG ATG AAA GGT GGG ATA CGA AAA GAC CGA AGA GGA GGG
TTC TCG ACG GTC CGG ACG GCC GAG GCA TTT ACG ATG CTT CAC CCT TAC TAC TTT CCA CCC TAT GCT TTT CTG GCT TCT CCT CCC
   ORSEQBAK           ────→
     TG TTG AAA CAC AAG CGC CAG
ARG MET LEU LYS HIS LYS ARG GLN ARG ASP ASP GLY GLU GLY ARG GLY GLU VAL GLY SER *** ***
AGA ATG TTG AAA CAC AAG CGC CAG AGA GAT GAT GGG GAG GGC AGG GGT GAA GTG GGG TCT TGA TAA TCAGGTCAGAGTGACCTGAGCT
TCT TAC AAC TTT GTG TTC GCG GTC TCT CTA CTA CCC CTC CCG TCC CCA CTT CAC CCC AGA ACT ATT AGTCCAGTCTCACTGGACTCGA

                          ←────
              EcoRI    M13FOR
                     TGACCGGCAGCAAAATG
AAAATAACACATTCAGAATTCACTGGCCGTCGTTTTAC
TTTTATTGTGTAAGTCTTAAGTGACCGGCAGCAAAATG
```

FIG. 2. DNA sequence of the N-terminal display vector pDM12 from M13BAK to M13FOR priming sites. Sequencing primer sites are shown above the DNA sequence. The pelB leader amino acid sequence is shown in italics. Restriction enzyme cloning sites are underlined and the 38-bp HRE DNA sequence is in bold, italicized type. The estrogen receptor DBD is separated from the cloning sites by a Gly_4 Ser linker fragment. *Sfi*I–*Not*I cleavage results in removal of the "stuffer" fragment; an additional *Pst*I digestion can be performed to improve the efficiency of subsequent library construction.

tion. The display of human leptin through the bacteriophage coat as a N-terminal fusion in pDM12 was confirmed by ELISA, which demonstrated significant binding of phage to polyclonal anti-leptin rabbit IgG in culture supernatant ($OD_{450\ nm}$ values of 0.1–0.3 above background). Sequencing of four ELISA-positive clones confirmed the presence of human leptin in frame with the leader and the DBD. In initial experiments, phage particles prepared from bacterial cultures containing human leptin cloned in frame with the DBD in the vector lacking the HRE site, pDM15, could not be bound by the anti-leptin rabbit IgG in ELISA. $OD_{450\ nm}$ values of 0.25 and 0.05 were obtained for pDM12–leptin and pDM15–leptin phage stocks, respectively. A background plate binding signal of 0.05 was observed for both phage stocks. However, in subsequent experiments performed with higher titer phage stocks prepared in an identical manner, binding of pDM15–leptin phage to antibody was observed, but always at a lower level than observed for pDM12–leptin phage (50–80% of pDM12–leptin phage ELISA signal). In view of this observation, wild-type M13K07 phage particles were tested for binding to recombinant estrogen receptor in ELISA, but binding was not observed (not shown).

To demonstrate the feasibility of C-terminal display, two different human cDNA fragments were fused to the C-terminus of the estrogen receptor DBD. The C-terminal cytoplasmic domain of human N-cadherin, a cell–cell adhesion protein, is recognized by a monoclonal anti-cadherin antibody which was raised against the C-terminal 24 amino acids of chicken N-cadherin (25). When cloned as a C-terminal fusion in pDM16 and packaged in bacteriophage, binding to anti-cadherin antibody was detected (Fig. 4). Binding was dependent on the presence of phage containing DBD–cadherin fusion DNA. The C-terminal domain of PCCα, a naturally biotinylated mitochondrial enzyme, is also biotinylated in *E. coli* (26). Particle binding to streptavidin was observed (Fig. 4), indicating the presence of biotinylated PCCα on the external surface of the bacteriophage. Again, binding to streptavidin was dependent on the presence of phage containing DBD–PCCα fusion DNA. The cDNA fragments displayed as C-terminal fusions contained 3′-UTR regions and poly(A) tail regions which appear not to be interfering with display on the external surface of the phage particle.

These initial experiments showed that both C- and N-terminal fusion proteins could protrude through the bacteriophage coat and interact with a specific ligand.

```
A
      NotI                             Oestrogen recept r DBD aa 176-282                                                              SalI
ALA ALA ALA GLY GLY GLY GLY SER MET GLU        VAL GLY SER Ter Ter
GCG GCC GCA GGG GGA GGA GGG TCC ATG GAA...   ......GTG GGG TCT TGA TAA TCAGGTCAGA GTGACCTGAG CTAAAATAAC ACATTCAG GTCGAC
CGC CGG CGT CCC CCT CCT CCC AGG TAC CTT...   ......CAC CCC AGA ACT ATT AGTCCAGTCT CACTGGACTC GATTTTATTG TGTAAGTC CAGCTG

                                      EcoRI
TTGGGTCAGT CTGACCGGGA CAAAGTTAAT GTAACCTC GAATTC
AACCCAGTCA GACTGGCCCT GTTTCAATTA CATTGGAG CTTAAG
B
      NotI                             Oestrogen receptor DBD aa 176-282                          EcoRI
ALA ALA ALA GLY GLY GLY GLY SER MET GLU        VAL GLY SER HIS HIS HIS HIS HIS HIS Ter Ter
GCG GCC GCA GGG GGA GGA GGG TCC ATG GAA..   ......GTG GGG TCT CAC CAC CAC CAC CAC CAC TGA TAA AGAA TTC
CGC CGG CGT CCC CCT CCT CCC AGG TAC CTT...  ......CAC CCC AGA GTG GTG GTG GTG GTG GTG ACT ATT TCTT AAG

C
HindIII
                                                MET LYS TYR LEU LEU PRO THR ALA ALA ALA GLY LEU LEU LEU LEU
AAGCTTGCAT GCAAATTCTA TTTCAAGGAG ACAGTCATAA ATG AAA TAC CTA TTG CCT ACG GCA GCC GCT GGA TTG TTA TTA CTC
TTCGAACGTA CGTTTAAGAT AAAGTTCCTC TGTCAGTATT TAC TTT ATG GAT AAC GGA TGC CGT CGG CGA CCT AAC AAT AAT GAG

                                   Oestrogen receptor DBD aa 176-282                        SfiI
ALA ALA GLN PRO ALA MET ALA GLU| MET GLU              VAL GLY SER | GLY GLY GLY GLY SER ALA GLN PRO ALA LEU
GCG GCC CAG CCG GCA ATG GCC GAG| ATG GAA.........  .........GTG GGG TCT | GGG GGA GGA GGG TCG GCC CAG CCG GCC CTC
CGC CGG GTC GGC CGT TAC CGG CTC| TAC CTT.........  .........CAC CCC AGA | CCC CCT CCT CCC AGC CGG GTC GGC CGG GAG

PstI        NotI                     SalI                                          EcoRI
LEU GLN LEU ALA ALA ALA Ter
CTG CAG CTG GCG GCC GCA TAACTGATTG AGTCGACTTG GGTCAGTCTG ACCGGGACAA AGTTAATGTA ACCTCGAATTC
GAC GTC GAC CGC CGG CGT ATTGACTAAC TCAGCTGAAC CCAGTCAGAC TGGCCCTGTT TCAATTACAT TGGAGCTTAAG
```

FIG. 3. (A) Partial DNA sequence of the N-terminal display vector pDM14. This vector contains two HRE sites separated by a *SalI* site. This vector was the template for construction of pDM16. (B) Partial DNA sequence of the N-terminal fusion vector pDM15. This vector was derived from pDM12 and has had the HRE site deleted; the six His tail is to allow purification of fusion proteins. Human leptin was cloned into this vector to demonstrate the importance of the HRE site for external surface display. (C) Partial DNA sequence of the C-terminal display vector pDM16.

This indicated that it may be possible to use the DBDX system for constructing and screening both oligo(dT) and random-primed cDNA libraries. However, it was possible that we were detecting partially formed, noninfective particles released from the cytoplasm by bacterial lysis during overnight growth. Repeated rounds of selection are required to allow enrichment of specific clones in phage display and "peptides on plasmids" systems. As DBD binding to phagemid DNA may have rendered the particles noninfective, it was necessary to confirm that we could enrich for infective, specific binding phage after elution from target. This was achieved using a N-terminally displayed cDNA library by selecting for a specific sequence known to be present in the library with an antibody.

Random-primed human lymphocyte cDNA was cloned in pDM12 as a N-terminal fusion library, producing ~10^8 clones, of which only about ~10^7 would be in frame with both the pelB and the DBD. Therefore, after amplification, most of the phage produced from this library would be nonbinding and have a significant growth advantage over in-frame clones. This library was analyzed for the presence of the desired protein by screening the library against a target coated onto a plastic tube. Selection against anti-human κ antibody resulted in an apparent enrichment between panning rounds 1 and 2 (Table 2). This suggested that binding clones were being selected and were still infective after standard high pH triethylamine elution (27). Screening 88 round 2 eluate clones revealed four ELISA-positive wells. Sequencing showed these to contain full-length human κ constant region, including a fragment of Jκ region DNA, which demonstrated that clones had been selected independently (not shown). This result compares favorably with a similar experiment performed elsewhere using a pJUN-FOS pIII lymphocyte cDNA library where four ELISA-positive wells were observed after four rounds of selection (28).

We further investigated the culture conditions to optimize the display of fusion protein per particle using a κ constant region clone isolated by screening the lymphocyte cDNA library. Standard pIII phage display libraries are normally amplified at 37°C overnight (27, 29) and we continued this practice for the initial lymphocyte library screening. However, having previously found that lowering the temperature of amplification for individual pIII fusion clones resulted in increased signal in ELISA, probably resulting from improved folding of the displayed protein, a similar effect was observed with the DBD fusion system. Lowering overnight incubation temperature to 30°C greatly enhanced binding to antibody, while there seemed to be a

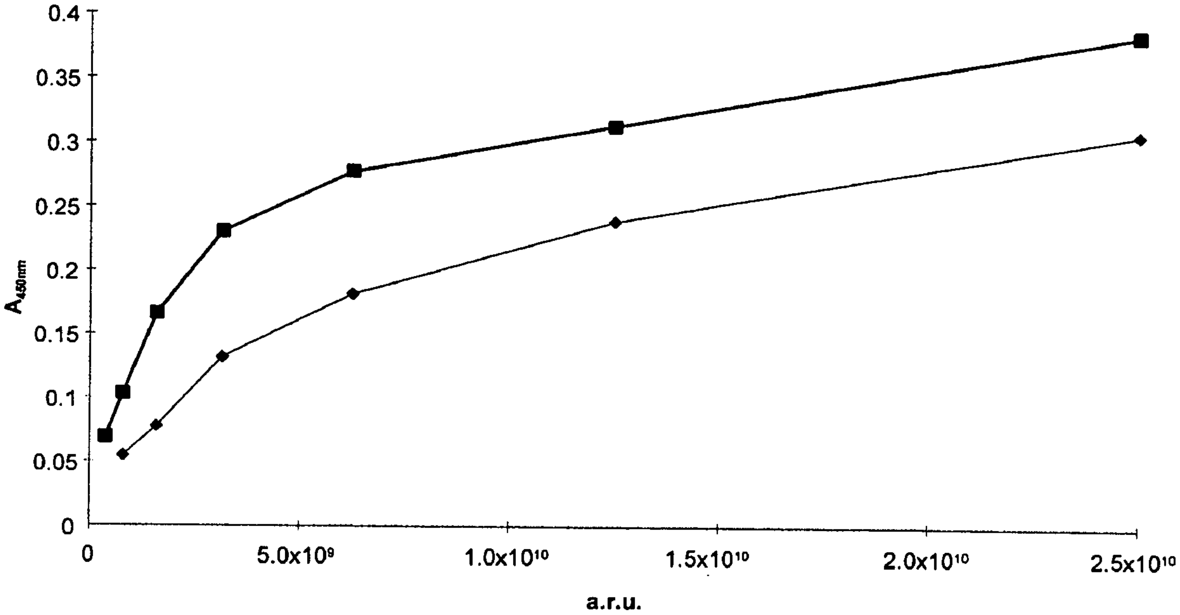

FIG. 4. Binding serial dilutions of C-terminally displayed N-cadherin (■) and PCCα (◆), to plates coated with anti-cadherin monoclonal antibody or streptavidin, respectively. Background $A_{450\text{ nm}}$ (M13K07) = 0.05. a.r.u., ampicillin resistance units.

less pronounced increase in binding with increasing IPTG concentration (Fig. 5).

DISCUSSION

It is essential for successful library screening that genotype and phenotype are linked and that this remains the case during the actual screening process without loss of genetic material associated with the binding protein. In addition, the binding moiety must be accessible to the target. The DBD display system described here meets these requirements. It demonstrates that proteins fused to a DNA-binding domain can be displayed externally, through a bacteriophage protein coat, while linked to the DNA encoding the displayed protein. We have also shown that proteins can be displayed as both N- and C-terminal fusions and that proteins with the desired binding properties can be selected from display libraries. Full-length estrogen receptor binding to HRE DNA in a vector has been shown to be highly pH dependent with an optimum between 7.5 and 8.0. At pH greater than 8.0 the receptor dissociates from the HRE sequence (18).

TABLE 2

Summary of Lymphocyte cDNA Library Selection

Selection round	Input	Output	% Recovery
1	4×10^{12}	10^5	2.5×10^{-6}
2	1.2×10^{12}	5×10^6	4×10^{-4}

Note. Input and output were measured as ampicillin resistance units (a.r.u.).

However, the requirement for an optimal HRE site to be present to allow external surface display has not been demonstrated, although its presence appears to enhance the number of fusion proteins displayed on average per particle. It maybe that under certain culture conditions, nonspecific, low-affinity association of the DBD with vector DNA may occur as has been demonstrated *in vitro* (18). Thus it may be possible for the DBD to bind with reduced affinity, to phagemid HRE-like sequences, especially during single-stranded DNA production when a relatively high DNA concentration will be reached in the bacterial cytoplasm and secondary structures formed in the single-stranded DNA.

The mechanism of DBD binding is more complex than the simple recognition of a target sequence.

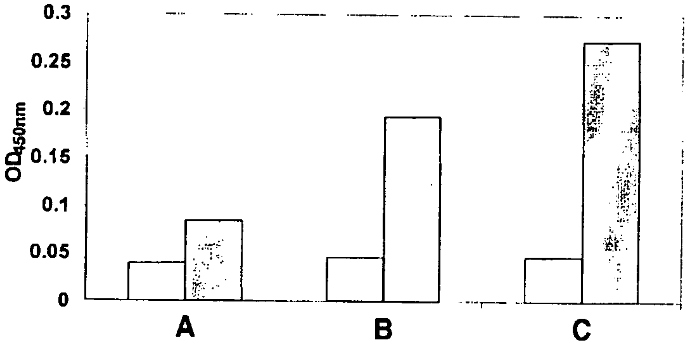

FIG. 5. Effect of temperature and IPTG concentration on ELISA signal for a N-terminal display fusion human κ constant region DBDX clone. IPTG concentrations of 20 μM (A), 50 μM (B), or 100 μM (C) were tested at 30°C (gray bars) or 37°C (open bars). Background $A_{450\text{ nm}}$(M13K07) = 0.03.

Flanking region sequences contribute significantly to the strength of binding (30, 31), and mutations within the palindrome are tolerated *in vivo* (31). Also, the DBD binds to the hexamer half-site sequence (TGACCT) with 66 nM affinity (32) and to double-stranded plasmid pGEM-1 DNA with 200 nM affinity (18). Indeed, this latter report showed that only three times the amount of estrogen receptor bound to HRE containing vector, compared with pGEM-1 vector alone *in vitro*. There are four TGACC sites in the pUC119 phagemid sequence which potentially could be bound by the DBD.

In order to display the fusion protein, the estrogen receptor DBD must associate with the target HRE sequence, or other DNA segment, in the cytoplasm of the host bacterium, displacing bacteriophage pV from the DNA. Such an association has already been demonstrated elsewhere (17). Interestingly, preliminary experiments have shown that the inclusion of zinc in the culture medium appears to increase the average number of fusion proteins displayed per particle (D. P. McGregor, unpublished). Zinc ions are an absolute requirement for high-affinity DNA binding *in vivo* and *in vitro* (32).

Constraints on the type and size of proteins displayed by DBD display are likely to result from interference in the phage particle assembly process. Once a particle has been extruded from the bacterium, it can be treated as any other phage particle derived from a standard phage display library. Whether disruption of the coat by the DBD reduces phage particle integrity is presently unknown, but such interference appears slight as completion of the phage life cycle and recovery of fusion protein DNA is unaffected.

The DBDX system differs from phage display and "peptide on plasmids" systems as it allows external surface display of a DBD fusion protein without using a bacteriophage coat protein fusion; also the same DBD can be used for both N- and C-terminal display. The system may offer some advantages over the established display methods. For example, during particle assembly there is no competition between wild-type phage coat proteins and recombinant coat protein fusions.

DBDX should be amenable to most of the uses already developed for bacterial-based display systems, such as Fab and single-chain antibody libraries (29, 33), with potentially a greater degree of versatility than current systems. We are currently investigating procedures to optimize the average number of DBD fusion proteins displayed per particle by varying the number and sites of HRE sequences together with changes in culture conditions. It has been shown *in vitro* that the presence of four HRE sites in a vector results in a corresponding increase in the amount of estrogen receptor bound compared with vectors containing one HRE site (18). Finally, it may be possible to integrate this system with bacteriophage coat protein fusions to create bifunctional phage particles.

ACKNOWLEDGMENTS

We thank Moira Johnson for sequencing and Nick Harries for critical review of the manuscript. We are indebted to the Scottish Executive Rural Affairs Department for support.

REFERENCES

1. Smith, G. P. (1985) Filamentous fusion phage: Novel expression vectors that display cloned antigens on the virion surface. *Science* **228**, 1315–1317.
2. Scott, J. K., and Smith, G. P. (1990) Searching for peptide ligands with an epitope library. *Science* **249**, 386–390.
3. Cwirla, S. E., Peters, E. A., Barrett, R. W., and Dower, W. J. (1990) Peptides on phage: A vast library of peptides for identifying ligands. *Proc. Natl. Acad. Sci. USA* **87**, 6378–6382.
4. Devlin, J. J., Panganiban, L. C., and Devlin, P. E. (1990) Random peptide libraries: A source of specific protein binding molecules. *Science* **249**, 404–406.
5. McCafferty, J., Griffiths, A. D., Winter, G., and Chiswell, D. J. (1990) Phage antibodies: Filamentous phage displaying antibody variable domains. *Nature* **348**, 552–554.
6. Barbas, C. F., III, Kang, A. S., Lerner, R. A., and Benkovic, S. J. (1991) Assembly of combinatorial antibody libraries on phage surfaces: The gene III site. *Proc. Natl. Acad. Sci. USA* **88**, 7978–7982.
7. Garrad, L. J., Yang, M., O'Connelll, M. P., Kelley, R. F., and Henner, D. J. (1991) F_{AB} assembly and enrichment in a monovalent phage display system. *BioTechnology* **9**, 1373–1377.
8. Lowman, H. B., and Wells, J. A. (1993) Affinity maturation of human growth hormone by monovalent phage display. *J. Mol. Biol.* **234**, 564–578.
9. Gram, H., Strittmatter, U., Lorenz, M., Gluck, D., and Zenke, G. (1993) Phage display as a rapid gene expression system: Production of bioactive cytokine-phage and generation of neutralizing monoclonal antibodies. *J. Immunol. Methods* **161**, 169–176.
10. Malik, P., Teryy, T. D., Gowda, L. R., Langara, A., Petukhov, S. A., Symmons, M. F., Welsh, L. C., Marvin, D. A., and Perham, R. N. (1996) Role of capsid structure and membrane protein processing in determining the size and copy number of peptides displayed on the major coat protein of filamentous bacteriophage. *J. Mol. Biol.* **260**, 9–21.
11. Gao, C., Mao, S., Chih-Hung, L. L., Wirsching, P., Lerner, R. A., and Janda, K. D. (1999) Making artificial antibodies: A format for phage display of combinatorial heterodimeric arrays. *Proc. Natl. Acad. Sci. USA* **96**, 6025–6030.
12. Jespers, L. S., Messens, J. H., De Keyser, A., Eeckhout, D., Van Den Brande, I., Gansemans, Y. G., Lauwereys, M. J., Vlasuk, G. P., and Stanssens, P. E. (1995) Surface expression and ligand-based selection of cDNAs fused to filamentous phage gene VI. *Bio/Technology* **13**, 378–382.
13. Cull, M. G., Miller, J. F., and Schatz, P. J. (1992). Screening for receptor ligands using large libraries of peptides linked to the C-terminus of the lac repressor. *Proc. Natl. Acad. Sci. USA* **89**, 1865–1869.
14. Schatz, P. J., Cull, M. G., Martin, E. L., and Gates, C. M. (1996) Screening of peptide libraries linked to lac repressor. *Methods Enzymol.* **267**, 171–191.

15. Greene, G. L., Gilna, P., Waterfield, M., Baker, A., Hort, Y., and Shine, J. (1986) Sequence and expression of human estrogen receptor complementary DNA. *Science* **231**, 1150–1154.
16. Mader, S., Chambon, P., and White, J. H. (1993) Defining a minimum estrogen receptor DNA binding domain. *Nucleic Acids Res.* **21**, 1125–1132.
17. Chusacultanachai, S., Glenn, K. A., Rodriguez, A. O., Read, E. K., Gardner, J. F., Katzenellenbogen, B. S., and Shapiro, D. J. (1999) Analysis of estrogen response element binding by genetically selected steroid receptor DNA binding domain mutants exhibiting altered specificity and enhanced affinity. *J. Biol. Chem.* **274**, 23591–23598.
18. Peale, F. V., Ludwig, L. B., Zain, S., Hilf, R., and Bambara, R. A. (1988) Properties of a high-affinity DNA binding site for estrogen receptor. *Proc. Natl. Acad. Sci. USA* **85**, 1038–1042.
19. Murdoch, F. E., Grunwald, K. A., and Gorski, J. (1991) Marked effects of salt on estrogen receptor binding to DNA: Biologically relevant discrimination between DNA sequences. *Biochemistry* **30**, 10838–10844.
20. Lannigan, D. A., and Notides, A. C. (1989) Estrogen receptor selectively binds the "coding strand" of an estrogen responsive element. *Proc. Natl. Acad. Sci. USA* **86**, 863–867.
21. Hoogenboom, H. R., Griffiths, A. D., Johnson, K. S., Chiswell, D. J., Hudson, P., and Winter, G. (1991) Multi-subunit proteins on the surface of filamentous phage: Methodologies for displaying antibody (Fab) heavy and light chains. *Nucleic Acids Res.* **19**, 4133–4137.
22. Higuchi, R., Krummel, B., and Saiki, R. K. (1988) A general method of *in vitro* preparation and specific mutagenesis of DNA fragments: Study of protein and DNA interactions. *Nucleic Acids Res.* **16**, 7351–7367.
23. Gong, D. W., Bi, S., Pratley, R. E., and Weintraub, B. D. (1996) Genomic structure and promoter analysis of the human obese gene. *J. Biol. Chem.* **271**, 3971–3974.
24. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual,. 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
25. Geiger, B., Volberg, T., Ginsberg, D., Bitzur, S., Sabanay, I., and Hynes, R. O. (1990) Broad spectrum pan-cadherin antibodies, reactive with the C-terminal 24 amino acid residues of N-cadherin. *J. Cell Sci.* **97**, 607–614.
26. Leon-Del-Rio, A., and Gravel, R. A. (1994) Sequence requirements for the biotinylation of carboxyl-terminal fragments of human propionyl-CoA carboxylase alpha subunit expressed in *Escherichia coli. J. Biol. Chem.* **269**, 22964–22968.
27. Marks, J. D., Hoogenboom, H. R., Bonnert, T. P., McCafferty, J., Griffiths, A. D., and Winter, G. (1991) By-passing immunization: Human antibodies from V-gene libraries displayed on phage. *J. Mol. Biol.* **222**, 581–597.
28. Light, J., Maki, R., and Assa-Munt, N. (1996) Expression cloning of cDNA by phage display selection. *Nucleic Acids Res.* **24**, 4367–4368.
29. Burton, D. R., Barbas, C. F., III, Persson, M. A. A., Koenig, S., Chanock, R. M., and Lerner, R. A. (1991) A large array of human monoclonal antibodies to type 1 human immunodeficiency virus from combinatorial libraries of asymptomatic seropositive individuals. *Proc. Natl. Acad. Sci. USA* **88**, 10134–10137.
30. Anolik, J. H., Klinge, C. M., Hilf, R., and Bambara, R. A. (1995) Cooperative binding of estrogen receptor to DNA depends on spacing of binding sites, flanking sequence, and ligand. *Biochemistry* **34**, 2511–2520.
31. Driscoll, M. D., Sathya, G., Muyan, M., Klinge, C. M., Hilf, R., and Bambara, R. A. (1998) Sequence requirements for estrogen receptor binding to estrogen response elements. *J. Biol. Chem.* **273**, 29321–29330.
32. Predki, P. F., and Sarkar, B. (1992) Effect of replacement of "zinc finger" zinc on estrogen receptor DNA interactions. *J. Biol. Chem.* **267**, 5842–5846.
33. Vaughan, T. J., Williams, A. J., Pritchard, K., Osbourn, J. K., Pope, A. R., Earnshaw, J. C., McCafferty, J., Hodits, R. A., Wilton, J., and Johnson, K. S. (1996) Human antibodies with sub-nanomolar affinities isolated from a large non-immunized phage display library. *Nature Biotechnol.* **14**, 309–314.

ANSWER 4 OF 15 CAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 1997:14393 CAPLUS
DOCUMENT NUMBER: 126:101229
TITLE: Selection of proteins and peptides from libraries displayed on filamentous bacteriophage
AUTHOR(S): **McGregor, Duncan**
CORPORATE SOURCE: Aberdeen Science and Technology Park, Scotgen Ltd, Campus 2, Aberdeen, UK
SOURCE: Mol. Biotechnol. (1996), 6(2), 155-162
CODEN: MLBOEO; ISSN: 1073-6085
PUBLISHER: Humana
DOCUMENT TYPE: Journal; General Review
LANGUAGE: English

AB A review with 54 refs. This article attempts to review recent developments in the rapidly developing field of phage display libraries. The current state of peptide, antibody, and cDNA libraries, as well as current and future applications of phage display libraries are discussed. The main focus of the article is on the methods for selecting binding ligands against targets in a variety of different formats. These include solid phase and in-soln. selection methods, and the strategies used to select for higher affinity, and binding ligands against impure and cellular target proteins.

REVIEW

Selection of Proteins and Peptides from Libraries Displayed on Filamentous Bacteriophage

Duncan McGregor

Reprinted with permission by the Publisher. This material is protected by copyright and cannot be further reproduced or stored electronically without publisher permission and payment of a royalty fee for each copy made. All rights reserved.

Abstract

This article attempts to review recent developments in the rapidly developing field of phage display libraries. The current state of peptide, antibody, and cDNA libraries, as well as current and future applications of phage display libraries are discussed. The main focus of the article is on the methods for selecting binding ligands against targets in a variety of different formats. These include solid phase and in solution selection methods, and the strategies used to select for higher affinity, and binding ligands against impure and cellular target proteins.

Index Entries: V_H, antibody heavy chain variable sequence region; V_L, antibody light chain variable sequence region; Fab, intact light chain (V_L and constant region domain [C_L]) associated with a truncated heavy chain composed of the V_H domain and the first constant region domain (C_H1); scFv, V_H and V_L domains linked as a single expression unit by a flexible linker; CDR, (complementarity determining region) region of hypervariable sequence within the antibody variable domain.

1. Introduction

Screening of vast synthetic peptide or biological peptide/protein libraries for therapeutic compounds allows the rapid isolation of active peptides or proteins that can then be used for therapy or as the basis for the rational drug design of synthetic chemical analogs. This article focuses on recent advances in *E. coli* filamentous bacteriophage gene III/VIII fusion protein peptide and protein libraries screening methodologies, with the purpose of describing in some detail the procedures for isolating novel binding entities, using bio-panning. Applications of phage display and library construction methods are reviewed extensively elsewhere *(1–4)*.

Interest in the display of peptides, antibodies, and other proteins on the surface of filamentous bacteriophage has resulted from the ability to link genetic material, packaged within the bacteriophage, with the selectable protein or peptide, displayed as a fusion protein on the surface *(5)*. Using the bio-panning affinity selection method, phage displaying antigenic determinants fused to the phage minor coat protein gene III can be quickly enriched relative to nonbinding phage *(6)*.

2. Libraries

2.1. Peptide Libraries

Peptide phage libraries are constructed using chemically synthesized degenerate oligonucleotides cloned into phage, or phagemid, vectors. Peptides are fused to either pIII or pVIII, and displayed at the amino termini of these proteins. Criteria for library construction have been reviewed elsewhere *(2,3)*. An interesting modification of standard peptide libraries uses a random peptide region that links human growth hormone (HGH) to pIII to identify protease substrates by immobilizing the library on a column of HGH receptor, then allowing specific proteases to release the phage displaying appropriate motifs *(7)*.

Most peptide ligands isolated from libraries have been found to represent linear epitopes from larger proteins, or have been isolated using targets known to recognize short peptides in vivo. This approach has been successful in epitope map-

Address to which all correspondence and reprint requests should be sent. Scotgen Ltd, Campus 2, Aberdeen Science and Technology Park, Aberdeen, Scotland.

Molecular Biotechnology ©1996 Humana Press Inc. All rights of any nature whatsoever reserved. 1073–6085/1996/6:2/155–162/$7.00

ping with antibodies, resulting in the isolation of peptides corresponding to the target protein sequence *(8)*, and peptide ligands for carbohydrate recognizing antibodies *(9)*. Discontinuous epitopes have been isolated from disulfide-constrained phage peptide libraries. It has also been noted that constraining peptide conformation can result in higher-affinity peptide-target interaction *(10)*. Finally, peptides isolated from longer insert libraries may be able to form a secondary structure not found in binding peptides from short insert libraries, resulting in completely different motifs being isolated *(11)*.

More recently, intracellular signal transduction domain ligands *(12)*, receptor antagonists *(13)*, tumor-suppressor protein-binding peptides *(14)* and immunogenic B-cell epitopes *(15)* have been identified.

2.2. Antibody Libraries

Potentially therapeutic human antibodies can be isolated from libraries with specificity to a wide range of self and foreign antigens. The affinity of these antibodies can then be increased by affinity maturation processes based on the selection of higher affinity variants by phage display (reviewed in refs. *1,2,4*).

Antibody libraries are constructed with V_H (antibody heavy chain variable sequence domain) and V_L (antibody light chain variable sequence domain) cDNA prepared from lymphocytes (peripheral blood, spleen, bone marrow, and so on) taken from a human or murine donor. PCR amplified antibody genes are then cloned into a phage(mid) vector in either a single-chain Fv (scFv) or two-domain Fab configuration *(16–18)*.

Two types of repertoire may be included in a library, depending on exposure to the antigen of interest, which can be termed "biased" (exposed) and "naive" (unexposed) *(19,20)*. Semi-synthetic repertoires (using degenerate oligonucleotides to encode V_H CDR3) can also be made to construct naive libraries or to select for catalytic properties *(21,22)*.

The size of a particular library is the crucial factor for the isolation of high-affinity antibodies. Biased libraries can be relatively small: The immune system has already had time to effect a response to the foreign antigen, and maturation of the antibody response will have occurred, resulting in the production of specific, high-affinity antibodies. Phage libraries constructed from such a source need only contain 10^6–10^7 individual members to allow the isolation of antibodies with affinities in the 10 n*M* range (typical of a secondary immune response). A good example of this approach is the work carried out by the Scripps group, who successfully isolated human antibodies to HIV, hepatitis B, cytomegalovirus (CMV), respiratory syncytial virus (RSV), and herpes simplex virus (HSV) *(19)*.

In order to isolate binding antibodies with reasonable affinities from naive libraries, very large libraries (>10^{10} clones) must be constructed. It proved necessary to utilize a bacterial recombination system *(Cre/lox)* to produce an Fab library of sufficient size (6.5×10^{10} clones) to isolate 10 n*M* affinity antibodies to the haptens NIP and fluorescein. Antibodies to a wide range of self and foreign antigens have been isolated from this library *(23)*.

Phage display has allowed the development of a number of methods for the in vitro affinity maturation of antibodies. These include Error prone PCR, in which point mutations are introduced into the antibody variable region gene by increasing the *Taq* polymerase error rate *(24)*, and the use of *E. coli* mutator strains to introduce random mutations, followed by selection, which has been found to yield higher affinity scFv fragments *(25)*. These two approaches are useful for obtaining single mutant antibodies with higher-affinity; however, constraints on library size prevent all possible double and triple mutants being screened, and suffer from the selection of point mutations that confer a growth advantage in *E. coli*. Recently, an approach developed by Affymax (Palo Alto, CA) has been devised which combines in vitro recombination with cassette mutagenesis of CDRs, potentially allowing the screening of all combinations of mutant and wild-type CDRs *(26)*. Alternatively, a codon based mutagenesis system can be used to create defined mutant CDR libraries, containing mixtures of wild-type and random sequence *(27)*.

Another method of affinity maturation is chain shuffling, a hierarchical approach, beginning with a known heavy chain and several unknown light chains. Panning of this library should reveal new light chains that confer higher affinity on the new antibody. Subsequently, a new library composed of the new light chain (plus VH CDR3) and unknown heavy chains is screened to increase affinity still further *(28)*.

The major drawback of phage antibodies isolated to date has been that, in most cases, the specific epitope that is reacted with is unknown. Standard panning methods select for antibodies that bind to the target, but these antibody fragments may not display any biological activity, e.g., viral neutralization. For example, work carried out using human anti-RSV antibodies demonstrates the problem: Of the three antibodies found to bind to the viral coat F glycoprotein, only one was effective in neutralizing the virus *(29)*. However, it has proved possible to direct the selection process to enrich for antibodies directed against neutralizing epitopes by the use of competitive elution techniques: competition between phage and well-characterized monoclonal antibodies *(30)*. Alternatively, selection can be directed by imprinting, a process that transfers the specificities of a murine monoclonal antibody to a novel human antibody by creating a chimeric library of the murine heavy chain and random human light chains, then selecting for binding antibodies and repeating the process with the new light chain and a library of human heavy chains in a procedure is analogous to chain shuffling *(31)*.

2.3. Other Protein Libraries

Phage display is being developed in two areas for nonantibody protein libraries. First, to increase the affinity and alter the specificity of particular proteins, and second, the isolation of previously uncharacterized proteins from cDNA libraries. Several proteins have recently been displayed on bacteriophage as geneIII fusions, including HGH, trypsin, plasminogen activator inhibitor, the high-affinity IgE receptor, ciliary neurotrophic factor, and others *(32–36)*. Most work to date has simply demonstrated that the displayed protein is functional, rather than attempting to define mutants with altered affinities or specificities. However, the display of IL3 on phage showed that an IL3-dependent cell line could be stimulated by the cytokine phage itself. This suggests that screening for biologically active variants will be possible using phage displayed proteins *(37)*.

Recently two successful methods for cDNA-phage library construction were described: A cDNA library was expressed as C-terminal fusion proteins to pVI, and serine protease inhibitors were isolated by standard panning techniques *(38)*. The second method is based on the co-expression of Fos-cDNA and Jun-pIII fusions, relying on the Jun-Fos leucine zipper dimerization to anchor the cDNA product to pIII *(39)*.

A final example of the display of nonantibody proteins is the isolation of bacterial receptor ligand-binding domains from fragmented DNA libraries *(40)*.

3. Enrichment Strategies and Improved Panning Methods for Selection of Phage

Bio-panning allows phage to associate with the target (either bound on a support, or in solution) over a period of time, followed by washing to remove nonspecifically bound phage, then elution of the remaining particles, and infection of *E. coli* to monitor the recovery. Amplification of the eluted phage, followed by repetition of the selection process, allows enrichment of specific-binding clones.

Solid support panning methods are similar to standard ELISA assays. A number of different materials have been used for coating target, including ELISA plates or ELISA tubes (e.g., NUNC [Paisley, Scotland] Immunotubes), or cross-linking to agarose *(17,20)*. The choice of support can affect the success of the panning procedure, depending on the target molecule concentration, and the displayed ligand affinity. Following coating, nonspecific sites on the matrix are blocked with either bovine serum albumin (BSA) or low fat milk powder in phosphate or Tris-buffered saline. Phage are then incubated with the target, in the presence of blocking agent, and then washed with buffer alone or supplemented with detergent (e.g., Tween 20) to remove nonspecific phage. The washing stringency can be varied by changing the number of washes for a particular round of panning to allow selection of high off-rate variants. Bound phage are then eluted from the support with either low pH glycine *(17,18)*, or high pH triethylamine *(20)*, neutralized with Tris buffer and allowed to infect log-phase male *E. coli*. A number of groups have found that bound phage can be efficiently eluted by incubating directly with *E. coli*. Tween 20 can also be included with the phage and the blocking solution. This can reduce nonspecific phage recovery (at 0.05% final concentration), but it can also reduce the recovery of specific phage; therefore, when panning against target on solid supports, including Tween 20, in the second and subsequent rounds of panning is preferred (unpublished observations).

Solution methods for selecting binding phage rely on the target molecule being tagged, normally with biotin, to allow the target/phage complexes to be separated from free phage onto a streptavidin-coated surface (either coated on an ELISA plate, or on magnetic beads) *(6,41)*. An aliquot of the phage library is mixed with the biotinylated target in the presence of blocking agent for a period of time that ensures equilibrium has been reached (normally, a few hours to overnight). Bound phage are removed from solution by streptavidin and the supernatant discarded. Washing and elution are then carried out as above. Thus, in this method there is a two-phase process to recovering bound phage: target-phage binding, followed by streptavidin-target binding. It has been suggested that this process increases the recovery of low-affinity ligands, because of multivalent interactions when a high-target molecule concentration is used *(42)*.

Because of the low numbers of a particular phage sequence in a primary phage population, the first round of selection is crucial to the successful isolation of binding clones: If 10^{11} phage, amplified from a library of 10^8 clones, are used for the first round of panning, there will be on average 1000 copies of each individual clone from the starting library. If only one binding clone is present in the library, and, for example, 1% of the specific binding phage are recovered, then only 10 copies will be recovered in the eluate. Following amplification of the recovered clones, enrichment in subsequent rounds of panning will allow recovery of increased numbers of phage, and an increase in the total number of phage recovered should be observed. The enrichment of binding phage between rounds of panning can also be monitored by performing a polyclonal ELISA with the amplified phage stocks from each round of panning. This is important, since it has been noted that the binding phage titer can drop if too many rounds of selection are performed *(14)*.

As the percentage of specific-binding phage recovered under particular panning conditions is approximately related to affinity and the concentration of target *(43)*, the low number of binding phage recovered in the first round of panning may be a concern when isolating ligands in the μ*M* range, such as antibodies from naive repertoires *(20,21)* or most peptides *(2,3)*. The situation is exacerbated when the target is a large macromolecule coated onto a solid support, resulting in a concentration of coated target well below the affinity constant of the phage-displayed ligand *(42)*.

In an experimental solid-phase panning system, adequate recovery of low-affinity ligands was achieved by performing a number of sequential panning steps in ELISA plate wells, or by increasing the phage concentration (in a reduced volume) in Immunotubes. In this experiment, the authors noted that the recovery of nonspecific phage was much lower in the ELISA plate format *(44)*.

Alternatively, vector systems which allow multivalent display of ligands can be used to allow avidity effects to encourage the recovery of low-affinity phage. Also, the *E. coli* strain used to produce the library may affect the number of correctly folded fusion-proteins displayed per phage: phage-antibodies produced as fusions to the C-terminal domain of pIII (i.e., a monovalent display system) in the *E. coli* TOP10F' strain (Invitrogen, de Schelp, Netherlands) have been shown to give higher ELISA signals than when produced in the widely used *E. coli* strain XL1-Blue (Stratagene, Cambridge, UK) *(45)*. Production of initial libraries in the former strain may increase the recovery of specific-binding phage during the first round of panning. In addition, the *E. coli* strain TG1 seems able to provide higher recoveries of infectious units, for the same input number, than other male *E. coli* strains (K. Johnson, personal communication). Increasing the number of phage incubated with target should also improve recovery in the first round of panning. This approach can be aided by the co-expression of the *E. coli* chaperonins, *Gro*EL and *Gro*ES, which can result in a 100-fold increase in initial library titers, thereby allowing more phage to be incubated with target *(46)*.

Panning methods can be subjected to numerous modifications for selective enrichment of high-affinity ligands. The simplest methods involve reducing the concentration of target molecules to a level that is below the affinity constant sought for the phage-displayed ligand. This approach has been used successfully in solution phase systems, demonstrating that phage can be selected according to their affinities, despite the possibility of multivalent interactions *(42,43)*. Another solution approach to discriminate between phage, is to select for phage-scFv antibodies with a slow off rate by initially incubating phage with a small amount of biotinylated antigen, followed by the addition of excess unbiotinylated antigen and following dissociation over time. This led to the selection of a mutant scFv with a >10-fold higher affinity for antigen than the parent antibody *(41)*.

Selective enrichment for specific binding phage can be achieved when screening libraries containing a protease cleavage site, such as trypsin or Factor Xa, at the fusion protein-geneIII junction *(45,47)*. In theory, treatment with a protease, following washing, should lead to the recovery of only those phage interacting with target via the fusion protein, rather than those that are nonspecifically bound through phage protein-target/coating matrix interaction. It has been shown for β-lactamase displayed on geneIII, that incubation with Factor Xa resulted in a 50-fold enrichment for specific phage over nonbinding phage in comparison with normal elution methods *(47)*. The success of this approach depends on equal accessibility of the cleavage site independent of the particular protein displayed. A similar approach can be applied when screening disulfide bridge-containing proteins or peptides by using DTT under conditions that have been shown not to affect phage infectivity to elute binding phage *(48)*. This was explained by Kremser and Rasched, who showed that the pIII disulfide bonds are internalized *(49)*. In a recent development, *E. coli* infection was linked to antigen recognition, allowing isolation of anti-HIV gp120 antibodies *(50)*. This was achieved by linking gp120 to geneIII and panning this with a naive Fab library produced with geneIII deficient M13. Phage bound to the gp120-geneIII fusion were then capable of infecting *E. coli*. When the same library was produced with normal M13, and panned against gp120, no binding clones were found.

As mentioned above, it is possible to ensure that phage are directed against an epitope of interest by competitively eluting binding phage with an antibody directed against a specific epitope *(30)*. This method allows the enrichment of specific binding antibodies or peptides from impure antigen sources. Alternatively, competitive elution, with antigen in a particular conformation, can be used to enrich for antibodies directed against epitopes exposed in antigen in an altered state. This was demonstrated in the isolation of antibodies specific for the cleaved C5a peptide, which did not bind to the intact C5 complement protein *(51)*.

It remains difficult to screen for ligands against many target molecules where the molecule is uncharacterized, or when purified to homogeneity, the native conformation of the molecule is lost, resulting in the exposure of areas of the molecule not normally seen in vivo, as in the case of membrane-bound proteins. However, a number of approaches have been developed through which ligands of unpurified targets can be isolated. Thyroid peroxidase (TPO), a 107 kDa protein expressed on the surface of thyroid cells, was expressed in Chinese hamster ovary (CHO) cells, and a phage-antibody library panned against the cells expressing the protein. Preincubating the phage antibody library with normal CHO cells overnight, and spinning down the cells to remove phage-antibodies to CHO cell surface markers, resulted in the isolation of anti-TPO specific autoantibodies *(52)*. This preabsorption technique has also been used in the isolation of anti-red blood cell antibodies, a situation where it was possible to detect positive clones by agglutination rather than ELISA (53). Finally, fluorescence-activated cell sorting has recently been applied to isolate phage-antibodies to particular cell types directly from peripheral blood leukocytes *(54)*.

4. Future Directions

It can be anticipated that advances in library construction techniques will allow the production of even larger libraries than can be made at present. These libraries, in conjunction with more efficient affinity maturation processes and novel selection methods which have recently begun to emerge, will ensure that phage display will play a key role in the isolation and improvement of lead therapeutic and diagnostic compounds for the pharmaceutical industry. Finally, phage display is likely to find a role in other areas not currently exploiting combinatorial library technology, such as in the improvement of toxic chemical degradation enzymes for bioremediation.

Acknowledgment

I would like to thank Dr. John Adair (Scotgen Biopharmaceuticals) for reviewing this manuscript and for many helpful discussions.

References

1. Burton, D. R. and Barbas, C. F., III (1994) Human antibodies from combinatorial libraries [Review], in *Advances In Immunology* vol. 41 (Dixon, F. J., ed.), Academic, Orlando, FL, pp. 191–280.
2. Clackson, T. and Wells, J. A. (1994) *In vitro* selection from protein and peptide libraries [Review]. *TIBTECH* **12,** 173–183.
3. Gallop, M. A., Barret, R. W., Dower, W. J., Fodor, S. P. A., and Gordon, E. M. (1994) Applications of combinatorial technologies to drug discovery. 1. Background and peptide combinatorial libraries [Review]. *J. Medicinal. Chem.* **37,** 1233–1251.
4. Winter, G., Griffiths, A. D., Hawkins, R. E., and Hoogenboom, H. R. (1994) Making antibodies by phage display technology [Review]. *Annu. Rev. Immunol.* 433–455.
5. Smith, G. P. (1985) Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* **228,** 1315–1317.
6. Parmley, S. F. and Smith, G. P. (1988) Antibody selectable filamentous fd phage vectors: affinity purification of target genes. *Gene* **73,** 305–318.
7. Matthews, D. J. and Wells, J. A. (1993) Substrate phage: selection of protease substrates by monovalent phage display. *Science* **260,** 1113–1117.
8. Jellis, C. L., Cradick, T. J., Rennert, P., Salinas, P., Boyd, J., Amirault, T., and Gray, G. S. (1993) Defining critical residues in the epitope for a HIV-neutralizing monoclonal antibody using phage display and peptide array technologies. *Gene* **137,** 63–68.
9. Hoess, R., Brinkmann, U., Handel, T., and Pastan, I. (1993) Identification of a peptide which binds to the carbohydrate-specific monoclonal antibody B3. *Gene* **128,** 43–49.
10. O'Neil, K. T., Hoess, R. H., Jackson, S. A., Ramachandran, N. S., Mousa, S. A., and Degrado, W. F. (1992) Identification of novel peptide antagonists for GPIIb/IIIa from a conformationally constrained phage peptide library. *Proteins* **14,** 509–515.
11. Kay, B. K., Adey, N. B., He, Y. S., Manfredi, J. P., Mataragnon, A. H., and Fowlkes, D. M. (1993) An M13 phage library displaying random 38-amino acid peptides as a source of novel sequences with affinity to selected targets. *Gene* **128,** 59–65.
12. Rickles, R. J., Botfield, M. C., Weng, Z., Taylor, J. A., Green, O. M., Brugge, J. S., and Zoller, M. J. (1994) Identification of Src, Fyn, Lyn, PI3K and Abl SH3 domain ligands using phage display libraries. *EMBO J.* **13,** 5598–5604.
13. Doorbar, J. and Winter, G. (1994) Isolation of a peptide antagonist to the thrombin receptor using phage display. *J. Mol. Biol.* **244,** 361–369.
14. Daniels, D. A. and Lane, D. P. (1994) The characterisation of p53 binding phage isolated from phage peptide display libraries. *J. Mol. Biol.* **243,** 639–652.
15. Motti, C., Nuzzo, M., Meola, A., Galfre, G., Felici, F., Cortese, R., Nicosia, A., and Monaci, P. (1994) Recognition by human sera and immunogenicity of HBsAg mimotopes selected from an M13 phage display library. *Gene* **146,** 191–198.
16. McCafferty, J., Griffiths, A. D., Winter, G., and Chiswell, D. J. (1990) Phage antibodies: filamentous phage displaying antibody variable domains. *Nature* **348,** 552–554.
17. Barbas, C. F., III, Kang, A. S., Lerner, R. A., and Benkovic, S. J. (1991) Assembly of combinatorial antibody libraries on phage surfaces: the gene III site. *Proc. Natl. Acad. Sci. USA* **88,** 7978–7982.
18. Garrad, L. J., Yang, M., O'Connelll, M. P., Kelley, R. F., and Henner, D. J. (1991) F_{AB} Assembly and enrichment in a monovalent phage display system. *BioTechnology* **9,** 1373–1377.

19. Williamson, R. A., Burioni, R., Sanna, P. P., Partridge, L. J., Barbas, C. F., III and Burton, D. R. (1993) Human monoclonal antibodies against a plethora of viral pathogens from single combinatorial libraries. *Proc. Natl. Acad. Sci. USA* **90,** 4141–4145.
20. Marks, J. D., Hoogenboom, H. R., Bonnert, T. P., McCafferty, J., Griffiths, A. D., and Winter, G. (1991) By-passing immunization. Human antibodies from V-gene libraries displayed on phage. *J. Mol. Biol.* **222,** 581–597.
21. Nissim, A., Hoogenboom, H. R., Tomlinson, I. M., Flynn, G., Midgley, C., Lane, D., and Winter, G. (1994) Antibody fragments from a 'single pot' phage display library as immunochemical reagents. *EMBO J.* **13,** 692–698.
22. Janda, K. D., Lo, C. H., Li, T., Barbas, C. F., III, Wirsching, P., and Lerner, R. A. (1994) Direct selection for a catalytic mechanism from combinatorial antibody libraries. *Proc. Natl. Acad. Sci. USA* **91,** 2532–2536.
23. Griffiths, A. D., Williams, S. C., Hartley, O., Tomlinson, I. M., Waterhouse, P., Crosby, W. L., Kontermann, R. E., Jones, P. T., Low, N. M., Allison, T. J., Prospero, T. D., Hoogenboom, H. R., Nissim, A., Cox, J. P. L., Harrison, J. L., Zaccolo, M., Gherardi, E., and Winter, G. (1994) Isolation of high affinity human antibodies directly from large synthetic repertoires. *EMBO J.* **13,** 3245–3260.
24. Gram, H., Marconi, L. A., Barbas, C. F., III, Collet, T. A., Lerner, R. A., and Kang, A. S. (1992) *In vitro* selection and affinity maturation of antibodies from a naive combinatorial immunoglobulin library. *Proc. Natl. Acad. Sci. USA* **89,** 3576–3580.
25. Irving, R. and Hudson, P. (1995) Superseding hybridoma technology with phage display libraries, in *Monoclonal Antibodies: The Second Generation* (Zola, H., ed.), BIOS Scientific, Oxford, UK.
26. Crameri, A. and Stemmer, W. P. C. (1995) Combinatorial multiple cassette mutagenesis creates all the permutations of mutant and wild-type sequences. *BioTechniques* **18,** 194–196.
27. Glaser, S. M., Yelton, D. E., and Huse, W. D. (1992) Antibody engineering by codon-based mutagenesis in a filamentous phage vector system. *J. Immunol.* **149,** 3903–3913.
28. Marks, J. D., Griffiths, A. D., Malmqvist, M., Clackson, T. P., Bye, J. M., and Winter, G. (1992) By-passing immunisation: building high affinity human antibodies by chain shuffling. *Bio/Technology* **10,** 779–783.
29. Barbas, C. F., III, Crowe, J. E., Jr., Cababa, D., Jones, T., Zebedee, S. L., Murphy, B. R., Chanock, R. M., and Burton, D. R. (1992) Human monoclonal Fab fragments derived from a combinatorial library bind to respiratory syncitial virus F glycoprotein and neutralise infectivity. *Proc. Natl. Acad. Sci. USA* **89,** 10,164–10,168.
30. Meulemans, E. V., Slobbe, R., Wasterval, P., Ramaekers, F. C., and van Eys, G. J. (1994) Selection of phage-displayed antibodies specific for a cytoskeletal antigen by competitive elution with a monoclonal antibody. *J. Mol. Biol.* **244,** 353–360.
31. Jespers, L. S., Roberts, A., Mahler, S. M., Winter, G., and Hoogenboom, H. R. (1994) Guiding selection of human antibodies from phage display repertoires to a single epitope of an antigen. *Bio/Technology* **12,** 899–903.
32. Lowman, H. B. and Wells, J. A. (1993) Affinity maturation of human growth hormone by monovalent phage display. *J. Mol. Biol.* **234,** 564–578.
33. Corey, D. R., Shiau, A. K., Yang, Q., Janowski, B. A., and Craik, C. S. (1993) Trypsin display on the surface of bacteriophage. *Gene* **128,** 129–134.
34. Pannekoek, H., van Meijer, M., Schleef, R. R., Loskutoff, D. J., and Barbas, C. F., III. (1993) Functional display of human plasminogen-activator inhibitor 1. *Gene* **128,** 135–140.
35. Robertson, M. W. (1993) Phage and *Escherichia coli* expression of the human high affinity immunoglobulin E receptor alpha-subunit ectodomain. Domain localization of the IgE-binding site. *J. Biol. Chem.* **268,** 12,736–12,743.
36. Saggio, I., Gloaguen, I., and Laufer, R. (1995) Functional phage display of ciliary neurotrophic factor. *Gene* **152,** 35–39.
37. Gram, H., Strittmatter, U., Lorenz, M., Gluck, D., and Zenke, G. (1993) Phage display as a rapid gene expression system: production of bioactive cytokine-phage and generation of neutralizing monoclonal antibodies. *J. Immunol. Methods* **161,** 169–176.
38. Jespers, L. S., Messens, J. H., De Keyser, A., Eeckhout, D., Van Den Brande, I., Gansemans, Y. G., Lauwereys, M. J., Vlasuk, G. P., and Stanssens, P. E. (1995) Surface expression and ligand-based selection of cDNAs fused to filamentous phage gene VI. *Bio/Technology* **13,** 378–382.
39. Crameri, R., Jaussi, R., Menz, G., and Blaser, K. (1994) Display of expression products of cDNA libraries on phage surfaces. A versatile screening system for selective isolation of genes by specific gene-product/ligand interaction. *Eur. J. Biochem.* **226,** 53–58.
40. Jacobsen, K. and Frykberg, L. (1995) Cloning of ligand-binding domains of bacterial receptors by phage display. *BioTechniques* **18,** 878–885.
41. Hawkins, R. E., Russell, S. J., and Winter, G. (1992) Selection of phage antibodies by binding affinity. Mimicking affinity maturation. *J. Mol. Biol.* **226,** 889–896.
42. Barrett, R. W., Cwirla, S. E., Ackerman, M. S., Olson, A. M., Peters, E. A., and Dower, W. J. (1992) Selective enrichment and characterization of high affinity ligands from collections of random peptides on filamentous phage. *Anal. Biochem.* **204,** 357–364.

43. Dyson, M. R., Germaschewski, V., and Murray, K. (1995) Direct measurement via phage titre of the dissociation constants in solution of fusion phage-substrate complexes. *Nucl. Acids Res.* **23,** 1531–1535.
44. Kretzschmar, T., Zimmermann, C., and Geiser, M. (1995) Selection procedures for non-matured phage antibodies: a quantitative comparison and optimisation strategies. *Anal. Biochem.* **224,** 413–419.
45. Ørum, H., Anderson, P. S., Øster, A., Johansen, L. K., Riise, E., Bjørnvad, M., Svendson, I., and Engberg, J. (1993) Efficient methods for constructing comprehensive murine Fab libraries displayed on phage. *Nucl. Acids Res.* **21,** 4491–4498.
46. Söderlind, E., Lagerkvist, A. C. S., Dueñas, M., Malmborg, A., Ayala, M., Danielsson, L., and Borrebaeck, C. A. (1993) Chaperonin assisted phage display of antibody fragments on filamentous bacteriophage. *Bio/Technology* **11,** 503–507.
47. Soumillion, P., Jespers, L., Bouchet, M., Marchand-Brynaert, J., Winter, G., and Fastrez, J. (1994) Selection of beta-lactamase on filamentous bacteriophage by catalytic activity. *J. Mol. Biol.* **237,** 415–422.
48. Raconjac, J. R. and Model, P. (1994) The influence of the reducing agent dithiothreitol *in vitro* on infectivity of f1 particles. *Gene* **142,** 153.154.
49. Kremser, A. and Rasched, I. (1994) The adsorption protein of filamentous phage fd: assignment of its disulfide bridges and identification of the domain incorporated in the coat. *Biochemistry* **33,** 13,954–13,958.
50. Dueñas, M. and Borrebaeck, C. A. K. (1994) Clonal selection and amplification of phage displayed antibodies by linking antigen recognition and phage replication. *Bio/Technology* **12,** 999–1002.
51. Ames, R. S., Tornetta, M. A., Jones, C. S., and Tsui, P. (1994) Isolation of neutralizing anti-C5a monoclonal antibodies from a filamentous phage monovalent Fab display library. *J. Immunol.* **152,** 4572–4581.
52. Portolano, S., McLachlan, S. M., and Rapoport, B. (1993) High affinity, thyroid-specific human autoantibodies displayed on the surface of filamentous phage use V genes similar to other autoantibodies. *J. Immunol.* **151,** 2839–2851.
53. Marks, J. D., Ouwehand, W. H., Bye, J. M., Finnern, R., Gorrick, B. D., Voak, D., Thorpe, S. J., Hughes-Jones, N. C., and Winter, G. (1993) Human antibody fragments specific for human blood group antigens from a phage display library. *Bio/Technology* **11,** 1145–1149.
54. Kruif, J., Terstappen, L., Boel, E., and Logtenberg, T. (1995) Rapid selection of cell subpopulation specific human monoclonal antibodies from a synthetic phage antibody library. *Proc. Natl. Acad. Sci. USA* **92,** 3938–3942.

Bioconjugate Chem. **1997**, *8*, 481–488

Synthesis and Characterization of a Peptide Nucleic Acid Conjugated to a D-Peptide Analog of Insulin-like Growth Factor 1 for Increased Cellular Uptake

Soumitra Basu[†] and Eric Wickstrom*

Department of Microbiology and Immunology and Kimmel Cancer Center, Thomas Jefferson University, Philadelphia, Pennsylvania 19107. Received January 27, 1997[®]

DNA therapeutics show great potential for gene-specific, nontoxic therapy of a wide variety of diseases. The deoxyribose phosphate backbone of DNA has been modified in a number of ways to improve nuclease stability and cell membrane permeability. Recently, a new DNA derivative with an amide backbone instead of a deoxyribose phosphate backbone, peptide nucleic acid (PNA), has shown tremendous potential as an antisense agent. Although PNAs hybridize very strongly and specifically to RNA and DNA, they are taken up by cells very poorly, limiting their potential as nucleic acid binding agents. To improve cellular uptake of a PNA sequence, it was conjugated to a D-amino acid analog of insulin-like growth factor 1 (IGF1), which binds selectively to the cell surface receptor for insulin-like growth factor 1 (IGF1R). The IGF1 D-peptide analog was assembled on (4-methylbenzhydryl)amine resin, and then the PNA was extended as a continuation of the peptide. The conjugate and control sequences were radiolabeled with ^{14}C or fluorescently labeled with fluorescein isothiocyanate. Cellular uptake of the PNA–peptide conjugate, a control with two alanines in the peptide, and a control PNA without the peptide segment were studied in murine BALB/c 3T3 cells, which express low levels of murine IGF1R, in p6 cells, which are BALB/c 3T3 cells which overexpress a transfected human IGF1R gene, and in human Jurkat cells, which do not express IGF1R, as a negative control. The specific PNA–peptide conjugate displayed much higher uptake than the control PNA, but only in cells expressing IGF1R. This approach may allow cell-specific and tissue-specific application of PNAs as gene-regulating agents *in vivo*.

INTRODUCTION

Targeting oligonucleotides to a particular gene, or messenger RNA of the gene, to specifically inhibit the expression of that gene has developed into an attractive therapeutic strategy in recent years, especially for treating cancers and viral diseases (*1*–*3*). Novel oligonucleotide analogs have been synthesized to act as antisense/antigene agents, to improve the biological stability, solubility, cellular uptake, and ease of synthesis. One of the recent additions to this group of modified oligonucleotide analogs is the peptide nucleic acid (PNA)[1] (Figure 1) (*4*). In these compounds, the entire deoxyribose phosphate backbone has been replaced with a structurally homomorphous polyamide (peptide) backbone composed of (2-aminoethyl)glycine units, leaving the oligomer uncharged. This synthetic DNA mimic exhibits enhanced affinity and specificity for its complementary nucleic acid target sequence.

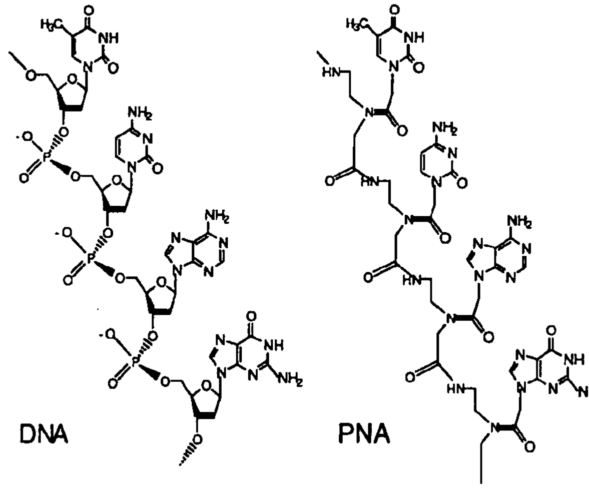

Figure 1. Structure of PNA and DNA. PNAs are peptide-based analogs of DNA in which the phosphate sugar backbone is replaced by (2-aminoethyl)glycine.

The PNAs have clear advantages over a variety of oligonucleotide analogs in several properties that are critical for antigene/antisense activity. Compared with other oligonucleotide derivatives, PNAs display the highest T_m values for duplexes formed with single-stranded DNA or RNA (*5*). PNAs are also resistant to both proteases and nucleases (*6*). Another major advantage is that PNAs can strand-invade duplex DNA, resulting in the formation of D-loops (*7*). This characteristic may make it possible to manipulate gene expression at the level of transcription. These complexes mediate the antigene/antisense effects of PNAs by steric hindrance

---

* Address correspondence to this author at the Department of Microbiology and Immunology, Thomas Jefferson University, 1025 Walnut St., Suite 420, Philadelphia, PA 19107 [telephone (215) 955-4578; fax (215) 955-4580; e-mail ewick@lac.jci.tju.edu].

[†] Present address: Department of Molecular Biophysics and Biochemistry, Yale University School of Medicine, New Haven, CT 06520.

[®] Abstract published in *Advance ACS Abstracts*, June 15, 1997.

[1] Abbreviations: Boc, *tert*-butyloxycarbonyl; Bzl, benzyl; DMEM, Dulbecco's modified Eagle's medium; FITC, fluorescein isothiocyanate; Fmoc, fluorenylmethoxycarbonyl; HPLC, high-performance liquid chromatography; IGF1, insulin-like growth factor; IGF1R, insulin-like growth factor 1 receptor; Mob, 4-methoxybenzyl; PNA, peptide nucleic acid; SDS, sodium dodecyl sulfate; SEM, standard error of the means; T_m, melting temperature; Z, benzyloxycarbonyl.

S1043-1802(97)00065-7 CCC: $14.00 © 1997 American Chemical Society

of enzyme complexes responsible for DNA transcription, cDNA synthesis, and RNA translation.

PNA activity as an antisense agent has been demonstrated *in vitro* and by microinjecting individual cells in culture (*8*). Microinjection of PNAs into cells was necessary because of poor cellular uptake (*9*) which was found to be 10 times less efficient than uptake of phosphorothioates in a variety of mammalian cells (*10*). One of the primary requirements for an oligonucleotide analog to be successful as an antigene/antisense agent is for it to be taken up by the cells in reasonable quantity so that it can reach its target in sufficient concentration. Since the PNAs suffer from poor cellular uptake, they have not been developed as an antigene/antisense therapeutic agent. To alleviate this situation, a strategy was developed to improve cellular uptake as well as to target the PNAs to specific cell types. A previous attempt was made to deliver PNA specifically through the blood–brain barrier (BBB) by binding a biotinylated PNA to streptavidin conjugated to a monoclonal antibody against transferrin receptor (*11*), to take advantage of the relatively high level of transferrin receptor at the BBB. The strategy was to have the PNAs cross the BBB via transferrin receptor mediated endocytosis. Though accumulation of the intravenously administered PNA–biotin–streptavidin–antibody inside the brain was 28-fold greater than accumulation of unmodified control PNA, no evidence was presented for cellular uptake of these large, complex conjugates.

Recent investigations have revealed that the insulin-like growth factor 1/insulin-like growth factor 1 receptor (IGF1/IGF1R) system plays major regulatory roles in development, cell cycle progression, and the early phase of tumorigenicity (*12*). Small peptides have been designed by molecular modeling as analogs of natural IGF1. The most effective peptide analog, JB3, D-Cys-Ser-Lys-Ala-Pro-Lys-Leu-Pro-Ala-Ala-Tyr-Cys, inhibits growth of certain cancer cell lines and competes with the natural ligand for binding to the IGF1R (*13*). Thus, we hypothesized that conjugation of the D-peptide analog with an antisense PNA against an effective target sequence of IGF1R mRNA (*14*) would provide cell-type specificity, and increased cellular uptake, by those cells overexpressing IGF1R. The rationale behind this strategy is the prediction that the peptide moiety would specifically bind the cell surface receptor, in this case IGF1R, to concentrate the conjugates on the specific cells, and then the conjugate would get taken up by receptor mediated endocytosis (*15*). In the case of PNAs, a study of cellular uptake demonstrated active endocytosis, with 13% of the internalized radiolabel localized to the nuclei after 8 h (*10*). No toxicity was evident over 24 h of observation. Once inside the cell, some fraction of internalized PNA may then interact with its target nucleic acid in the cytoplasm or nucleus.

The JB3 peptide has two Cys residues, one at each terminus, which are disulfide linked to form a loop with limited flexibility, favoring a conformation for binding to the receptor. The use of D-amino acids gave the peptide stability against cellular proteases. A reverse sequence was synthesized with respect to the normal L-amino acid sequence to account for the reversal of chirality (*13*). To reduce the complexity of the synthesis, a smaller version of JB3, called JB9, D-Cys-Ser-Lys-Cys, was selected for conjugation with the PNA. The peptide segment was synthesized automatically using standard Fmoc coupling, after which the PNA moiety was extended from the N terminus of the peptide by manual Boc coupling (*16*).

Cellular uptake of the PNA–peptide conjugate, a control with two D-Ala residues in the peptide in place of D-Ser-Lys, and a control PNA without a peptide adduct were studied in murine BALB/c 3T3 cells, which express low levels of murine IGF1R, in p6 cells, which are BALB/c 3T3 cells, which overexpress a transfected human IGF1R gene (*13*), and in human Jurkat cells, which do not express IGF1R (*17*), as a negative control. In cells expressing IGF1R, the specific PNA–peptide conjugate displayed significantly higher uptake than the control PNA or the control PNA–peptide. This approach may allow cell-specific and tissue-specific application of PNAs as gene-regulating agents *in vivo*.

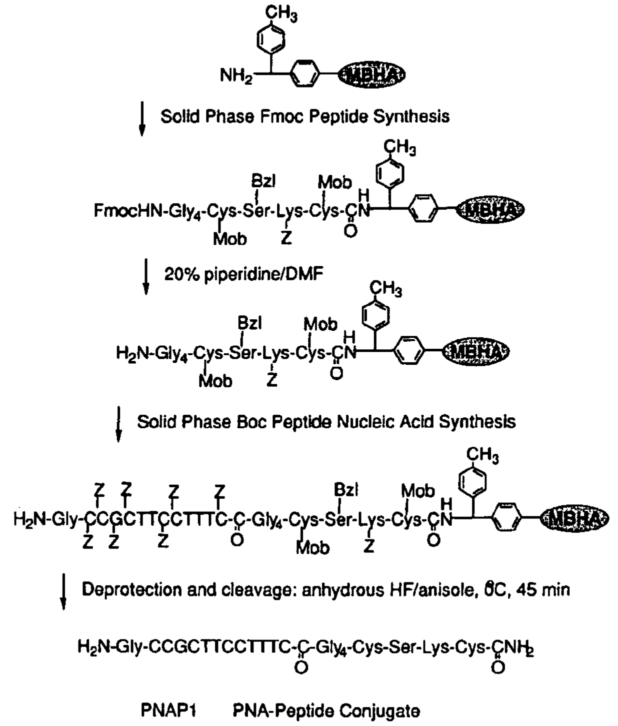
CH_3
NH_2
MBHA
Solid Phase Fmoc Peptide Synthesis
CH_3
Bzl
Mob
H
FmocHN-Gly_4-Cys-Ser-Lys-Cys-CN
Mob
Z
O
MBHA
20% piperidine/DMF
CH_3
Bzl
Mob
H
H_2N-Gly_4-Cys-Ser-Lys-Cys-CN
Mob
Z
O
MBHA
Solid Phase Boc Peptide Nucleic Acid Synthesis
CH_3
Z Z Z Z
Bzl
Mob
H
H_2N-Gly-CCGCTTCCTTTC-C-Gly_4-Cys-Ser-Lys-Cys-CN
Z Z Z O
Mob
Z
O
MBHA
Deprotection and cleavage: anhydrous HF/anisole, 0C, 45 min
H_2N-Gly-CCGCTTCCTTTC-C-Gly_4-Cys-Ser-Lys-Cys-CNH_2
O
O
PNAP1 PNA-Peptide Conjugate

Figure 2. Synthetic scheme for assembly of PNA–peptide conjugates.

MATERIALS AND METHODS

Assembly of PNA–Peptide Conjugates. The IGF1R-targeted PNA–peptide (PNAP1) synthesized was H-Gly-CCGCTTCCTTTC-Gly_4-Cys-Ser-Lys-Cys-NH_2 (Figure 2). The PNA dodecamer, CCGCTTCCTTTC, is complementary to nucleotides 2251–2262, codons 706–709, of human IGF1R mRNA, corresponding to the putative precursor processing site (*18*). The four glycines at the N terminus of the peptide adduct serve as a spacer between the peptide and the PNA moieties to minimize mutual interference, as these two segments by design have independent functions.

The peptide portion of the conjugate was synthesized automatically on *p*-methylbenzhydrylamine–HCl (1% divinyl polystyrene cross-linked) resin on an Applied Biosystems 331 peptide synthesizer at 0.10 mmol scale (250 mg) with a standard HBTU coupling protocol using Fmoc-D-Cys (Mob)-OH (Bachem Biosciences, Inc., King of Prussia, PA), Fmoc-D-Ser (Bzl)-OH (Bachem Biosciences), Fmoc-D-Lys (Z)-OH (Novabiochem, San Diego, CA), and Fmoc-Gly-OH (Applied Biosystems, Foster City, CA). After assembly of the complete peptide, the N-terminal Fmoc group was removed. Thus, the amino terminus became available for assembling the PNA chain as a continuation of the protected D-peptide segment. H_2N-Gly-Gly-Gly-Gly-D-Cys(Mob)-D-Ser(Bzl)-D-Lys (Z)-D-

Cys(Mob)–resin. The Boc-protected PNA monomers (PerSeptive Biosystems, Framingham, MA) were coupled manually essentially as described (*16*).

In short, 1.5 mL solutions of 0.10 M Boc-protected monomer (0.15 mmol) were preactivated for 2 min with 0.08 M HBTU and 0.20 M DIEA (base) in pyridine/DMF (1:1 v/v) and then coupled to 0.040 mmol (200 mg) of H-Gly-Gly-Gly-Gly-D-Cys(Mob)-D-Ser(Bzl)-D-Lys(Z)-D-Cys-(Mob)–resin for 20 min. Qualitative ninhydrin analysis was conducted on an aliquot of the growing Boc-PNA–peptide–resin to determine the presence of free amines (*19*). The resin was washed twice with 3 mL of pyridine for 2 min. The free amines were capped with a mixture of acetic anhydride/pyridine/CH_2Cl_2 (10:12:78, v/v/v) for 5 min. Any acetyl esters formed by the capping step were removed by treatment with 3 mL of piperidine/CH_2Cl_2 for 5 min. The resin was then washed three times with 3 mL of DMF/CH_2Cl_2 (1:1, v/v), followed by washing three times with 3 mL of neat CH_2Cl_2, three times, for 2 min per wash step. The terminal Boc group was removed by treating the resin twice with TFA/*m*-cresol (95:5, v/v) for 2 min. The PNA–peptide–resin was washed three times with DMF/CH_2Cl_2 (1:1, v/v) for 2 min each, followed by two washes with neat pyridine for 2 min each. The coupling/deprotection cycle with Boc-PNA monomers was repeated until the entire PNA sequence was assembled. Ninhydrin analysis was repeated and proved to be negative. The completed PNA–peptide conjugate was deprotected and cleaved from the resin with anhydrous HF/anisole at 0 °C for 45 min. A control PNA (PNA1) without a peptide, Gly-CCGCTTCCTTTC-$CONH_2$, and a peptide sequence control (PNAP2), Gly-CCGCTTCCTTTC-CONH-Gly_4-Cys-Ala-Ala-Cys-CO_2H, were custom synthesized by PerSeptive Biosystems.

Purification of PNA–Peptide Conjugates. The two cysteine side chains in the peptide were cyclized by dissolving the conjugate in 0.01 M $NaHCO_3$, pH 8.5, at 5.0 g/L, and stirring for 24 h at room temperature in the presence of atmospheric oxygen to allow formation of disulfide linkages. The lyophilized conjugate was assayed for free sulfhydryl groups by Ellman's reagent (DTNB assay) (*20*). The cyclized crude product was purified by HPLC on a 10 × 250 mm C_{18} Econosil column (Alltech Associates, Deerfield, IL) eluted over 40 min from 10% to 80% acetonitrile in water containing 0.1% TFA, at 4 mL/min, on a Waters 600 multisolvent delivery system coupled with a temperature controller maintaining the column at 50 °C and a Waters 486E variable-wavelength detector, monitoring eluent absorbance at 260 nm.

Characterization of PNA–Peptide Conjugates by Mass Spectroscopy. The conjugate was also characterized by MALDI-TOF mass spectroscopy (*21*) on a Hewlett-Packard 1700 LDI calibrated with a standard peptide mixture supplied by the manufacturer. Equal volumes of a 0.5 mM solution of the conjugate in 20% methanol, 80% water, and a solution of sinapinic acid (3,5-dimethoxy-4-hydroxycinnamic acid; Aldrich, Milwaukee, WI) were mixed well. Then, 1 μL of the solution was placed on the tip of the probe, and the sample was allowed to crystallize under vacuum and then analyzed to obtain the mass of the conjugate. A PNA1 sample was similarly analyzed.

Characterization of PNA–Peptide Conjugates by SDS–Polyacrylamide Gel Electrophoresis. The PNA–peptide conjugate was also characterized by SDS–polyacrylamide gel electrophoresis. The samples were electrophoresed on 4–20% acrylamide gradient gels (Bio-Rad, Hercules, CA), on a Bio-Rad Mini-Protean II Cell apparatus. Gel electrophoresis was conducted at 100 V, in Laemmli buffer, pH 8.3 (*22*), calibrated with low-range prestained protein molecular mass standards, 3–43 kDa (Life Technologies, Gaithersburg, MD). The gels were stained with Coomassie brilliant blue to visualize the bands.

Synthesis of [^{14}C]PNA–Peptide and [^{14}C]PNA Conjugates. PNA1, PNAP1, and PNAP2 were radioactively labeled by reductive methylation with [^{14}C]formaldehyde as described (*10*, *23*). Briefly, PNA and PNA–peptides (100 nmol) were dissolved in 0.1 mL of 0.2 M Na_2HPO_4, pH 7.5 (to minimize lysine methylation), with 500 nmol of [^{14}C]formaldehyde (NEC-039H, 40–60 Ci/mol, New England Nuclear, Boston, MA) and incubated for 2 h at room temperature with periodic vortexing. Sodium cyanoborohydride (0.1 mL of a fresh 100 mM solution) was then added to the mixture to reduce the Schiff base, and the incubation was continued for an additional 4 h with periodic vortexing. The [^{14}C]PNA derivatives were purified by gel filtration on NAP10 G-25 Sephadex columns (17-0854-01, Pharmacia Biotech Inc., Piscataway, NJ). Homogeneity was evaluated by TLC of a small aliquot on cellulose plates (1366061, Eastman Kodak, Rochester, NY) developed with *n*-butanol/glacial acetic acid/H_2O (4:1:5). Specific activities of the labeled PNAs were estimated by measuring concentrations from A_{260} in UV-absorbing TLC bands extracted with water and ^{14}C radioactivity using liquid scintillation counting at 75% counting efficiency. The specific activities of several preparations ranged from 5.3 to 11 Ci/mol.

Synthesis of PNA–Peptide–Fluorescein and PNA–Fluorescein Conjugates. Purified PNAP1 and PNA1 were fluoresceinated with a 20-fold excess of FITC in 0.2 M Na_2HPO_4 buffer, pH 8.5 (to minimize lysine fluoresceinylation), for 2 h with constant stirring. The products were purified from free FITC by preparative chromatography on a NAP10 G-25 Sephadex column and further purified on a G-50 Sephadex column (Pharmacia Biotech Inc.). Fluorescein conjugates were then analyzed by SDS–polyacrylamide gel electrophoresis as above.

Thermal Denaturation of PNA–DNA Duplexes. Thermal denaturation profiles of an equimolar mixture of the purified PNAP1 conjugate or PNA1 control, and their complementary DNA target, were performed in 10 mM Tris-HCl, pH 8.0, 100 mM NaCl, on a Cary 3E spectrophotometer equipped with a multicell holder and a temperature controller (Varian, Palo Alto, CA). The rate of increase of the temperature was 1 °C/min, from 10 to 90 °C. All T_m values were calculated from the first derivative of the melting curve, and the reported values are the average of three experiments.

Serum Stability of the PNA–Peptide Conjugate. An aliquot of PNAP1 conjugate was incubated in 10% fetal bovine serum/PBS for 12 and 24 h at 37 °C. After incubation, 1 volume of acetonitrile was added to each sample to precipitate serum proteins, which were pelleted by centrifugation (*24*). Free conjugate in the supernatant was evaporated *in vacuo*. It was then analyzed by reversed phase HPLC as described above.

Cellular Uptake Studies with [^{14}C]PNA and [^{14}C]-PNA–Peptide Conjugates. Cellular uptake studies with radiolabeled PNA derivatives were carried out as described previously (*10*). p6 cells, which are murine BALB/c 3T3 cells transfected with human IGF1R gene, and nontransfected murine BALB/c 3T3 cells were plated in DMEM with 10% FBS in 12.5 cm^2 flasks and then allowed to attach and grow for 2 days in a humidified cell incubator with 5% CO_2 at 37 °C, reaching 50–80% confluence, on the order of 10^6 cells per flask. The medium was then removed and replaced with fresh medium containing 1 μM of the various [^{14}C]PNA derivatives, prewarmed to 37 °C; background cells received medium with no [^{14}C]PNA derivatives. After incubation for various times, attached cells were washed directly in the flasks using four washes: once with fresh medium, once with PBS, once with 1.0 M NaCl/0.4 M NaOAc, pH 3.3, and once with PBS. This procedure had the advantage of removing both noninternalized oligomer and any dead cells, which can accumulate large amounts of labeled oligomer and thereby skew any uptake measurements (*10*). Human Jurkat cells, which lack IGF1R (*17*), were used as a negative control. They were grown in suspension in RPMI 1640 with 10% FBS in a humidified cell incubator with 5% CO_2 at 37 °C. The Jurkat cells were treated in the same manner as the attached cells, except that they were incubated and washed in centrifuge tubes, rather than on plates, because they grow in suspension.

After the final wash, the cells were lysed in 1 mL of 1% SDS in H_2O, and the samples were processed for total protein measurement and liquid scintillation counting as described before (*10*). Fifty microliters of the lysate was allocated for total protein measurement, and ^{14}C activity in the remaining lysate was measured by liquid scintillation counting, from which background counts were subtracted. Background samples from control lysates not treated with [^{14}C]PNA derivatives typically yielded 12–15 cpm, corresponding to <1 pmol even at 10 Ci/mol or 0.2 pmol at a typical specific activity of 40 Ci/mol. Thus, an experimental sample with twice the background counts per mimute would contain 0.2–1 pmol, depending on the specific activity, which may be considered the limit of detection.

Control samples of varying confluence were trypsinized following the last wash, resuspended in PBS for cell counting, and then lysed to obtain total protein values. This allowed the generation of a standard curve corresponding to total protein *vs* cell number. The curve was used to estimate cell counts from the total protein measurements of treated cells, and [^{14}C]PNA counts were used to calculate picomoles of oligomer using the specific activity of each labeled preparation. Values for cell number and picomoles of cell-associated oligomer were therefore obtained for each treated sample. Estimates of cell volume were obtained by microscopic evaluation of cell diameters using a micrometer and subsequent calculation of an average cell volumes, as described (*10*).

Cellular Uptake Studies with PNA–Fluorescein and PNA–Peptide–Fluorescein Conjugates. Cellular uptake studies with fluorescein conjugates of PNA1 and PNAP1 were conducted with p6 cells and Jurkat cells. P6 cells were plated on LAB TEK 8 well tissue culture Chamber Slides (Nunc, Naperville, IL) in DMEM with 10% FBS at a concentration of 20 000 cells/chamber. The cells were allowed to attach and grow for 24 h in a humidified cell incubator with 5% CO_2 at 37 °C. The attached cells were then washed with serum-free DMEM and PBS. The cells were incubated for 4 h at 37 °C with 1 μM of the fluoresceinated PNAP1 in PBS. Human Jurkat cells, the negative control, were treated in the same manner as the p6 cells, except that they were incubated and washed in centrifuge tubes rather than on plates, because they grow in suspension. At the conclusion of the incubation period the cells were washed with serum-free DMEM once followed by three washings with PBS. Then the cells were fixed with 1% paraformaldehyde/PBS for 1 h and washed once with buffer from an ANTI-FADE kit (Molecular Probes, Eugene, OR). The chambers were removed and excess liquid drained off. One drop of the ANTI-FADE reagent was sufficient to cover the cells on the slides, which were covered with coverslips and sealed. The fixed slides were examined on a Bio-Rad MRC-600 laser scanning confocal microscope, interfaced to a Zeiss Axiovert 100 with a Plan-Apo 63X 1.40NA oil-immersion lens, using fluorescence microscopy and phase contrast microscopy to observe cellular uptake of the fluorescent oligomers.

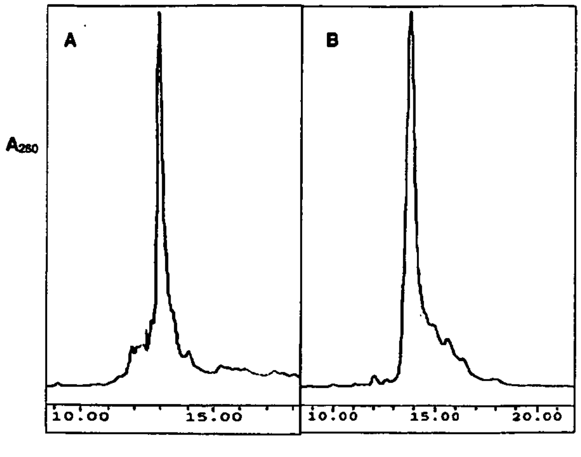

Figure 3. Analytical HPLC of purified, cyclized PNA–peptide conjugate PNAP1, on 4.6 × 250 mm Econosil C_{18} column eluted over 40 min from 12% to 76% acetonitrile in water containing 0.1% TFA, at 1 mL/min, at 50 °C: (A) following purification; (B) extracted from serum after 24 h of incubation. Eluent absorbance was monitored at 260 nm.

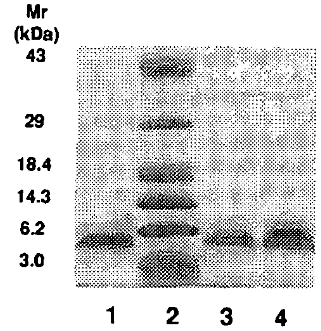

Figure 4. SDS–PAGE of the PNA–peptide conjugate: lanes 1 (0.5 nmol), 3 (0.75 nmol), and 4 (1.5 nmol) are the conjugate; lane 2 contains the molecular mass markers. The gel was stained with Coomassie brilliant blue. The conjugate (3.85 kDa) migrates between the 6.2 and the 3.0 kDa markers.

RESULTS

Assembly, Chromatography, and Mass Spectroscopy of the PNA–Peptide Conjugate. The PNA–peptide conjugate PNAP1, 5′-CCGCTTCCTTTC-3′-N-(Gly)$_4$-D-(Cys-Ser-Lys-Cys)-C, was synthesized on a solid phase support with a blend of automatic and manual syntheses (Figure 2). The conjugate was deprotected and cleaved from the solid support, yielding 140 mg (0.0364 mmol) of crude linear product, a 91% yield by mass, which could be an overestimate, due to the potential inclusion of nonpeptide impurities. Analytical HPLC displayed a main product peak including almost 90% of the total area under the peaks. Thus, the product of 91% crude yield with 90% full-length product in that material gives a maximum estimated yield of 82%. The linear PNAP1 was then cyclized and purified by reversed phase HPLC. Analytical HPLC after purification (Figure 3A) displayed a single peak. This is the first demonstration of a partially automated synthesis of a PNA–peptide

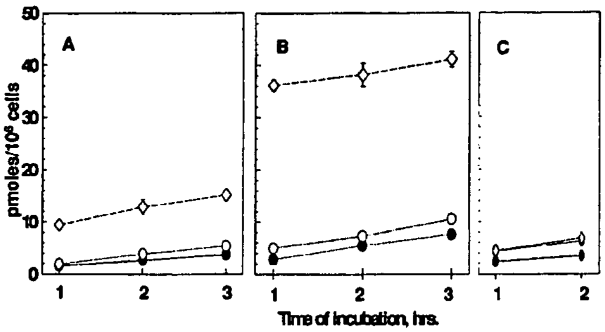

Figure 5. Cellular uptake of radiolabeled PNA and PNA–peptides by mammalian cells as a function of incubation time at 37 °C: (A) BALB/c 3T3 cells; (B) p6 cells; (C) Jurkat cells. Cells were incubated in the presence of 1 μM [^{14}C]PNA1 (●), [^{14}C]PNAP1 (◇), and [^{14}C]PNAP2 (○) for varying lengths of time and processed for measurement of cell-associated radioactivity. Data are presented in terms of picomoles of oligonucleotide per 10^6 cells. Each data point in (A) and (B) represents the mean ± SEM of three replicates. Each data point in (C) represents the mean ± variance of two replicates.

conjugate. Two different coupling chemistries were utilized: the more widely used Fmoc coupling for the peptide segment and Boc coupling for the PNA segment. The purified cyclized PNAP1 peak provided a MALDI-TOF mass spectrum with a prominent peak at 3854.5 amu, close to the calculated mass (3850.7 amu) of the intended PNAP1 conjugate. A negative result from the DTNB assay of purified, cyclized PNAP1 implied absence of free sulfhydryl groups, confirming the completeness of disulfide cyclization. A single peak in the HPLC chromatogram, as well as a single peak in the mass spectrum corresponding to the monomeric mass of PNAP1–peptide, indicated that the peptide portion was indeed cyclized, rather than dimerized or oligomerized. PNA1 was assembled, purified by HPLC, and analyzed by mass spectrum in the same fashion. PNAP2 from PerSeptive Biosystems was also homogeneous after cyclization and purification.

Electrophoretic Analysis of the PNA–Peptide Conjugate. To further characterize the conjugate, it was analyzed by denaturing SDS–polyacrylamide gel electrophoresis (*22*), used regularly for sizing of polypeptides. Previously, it has been possible to analyze complexes of PNAs hybridized to charged nucleic acids by nondenaturing gel electrophoresis (*5*). However, free PNAs have not been analyzed previously by gel electrophoresis, due to their lack of charge. The SDS-denatured conjugate migrated as predicted according to its molecular mass, compared to the standard peptide molecular mass markers (Figure 4). This is the first demonstration that PNA–peptide conjugates could be analyzed for their homogeneity and molecular mass by denaturing gel electrophoresis. Fluorescein conjugates were also analyzed by SDS–polyacrylamide gel electrophoresis and displayed single bands of lower mobility (not shown).

Thermal Denaturation of PNA–DNA Duplexes. The PNA–peptide conjugate includes a Gly_4 spacer between the dodecamer PNA segment and the tetrapeptide IGF1 analog. Despite the presence of the spacer section, it is plausible that the peptide might interfere with PNA hybridization to a complementary target. Therefore, a thermal denaturation study was undertaken to determine whether the peptide moiety might lower the T_m of the PNA/DNA hybrid. The two melting curves were identical, yielding a T_m of 60 ± 1 °C. These results imply that the peptide moiety on the C terminus of the PNA

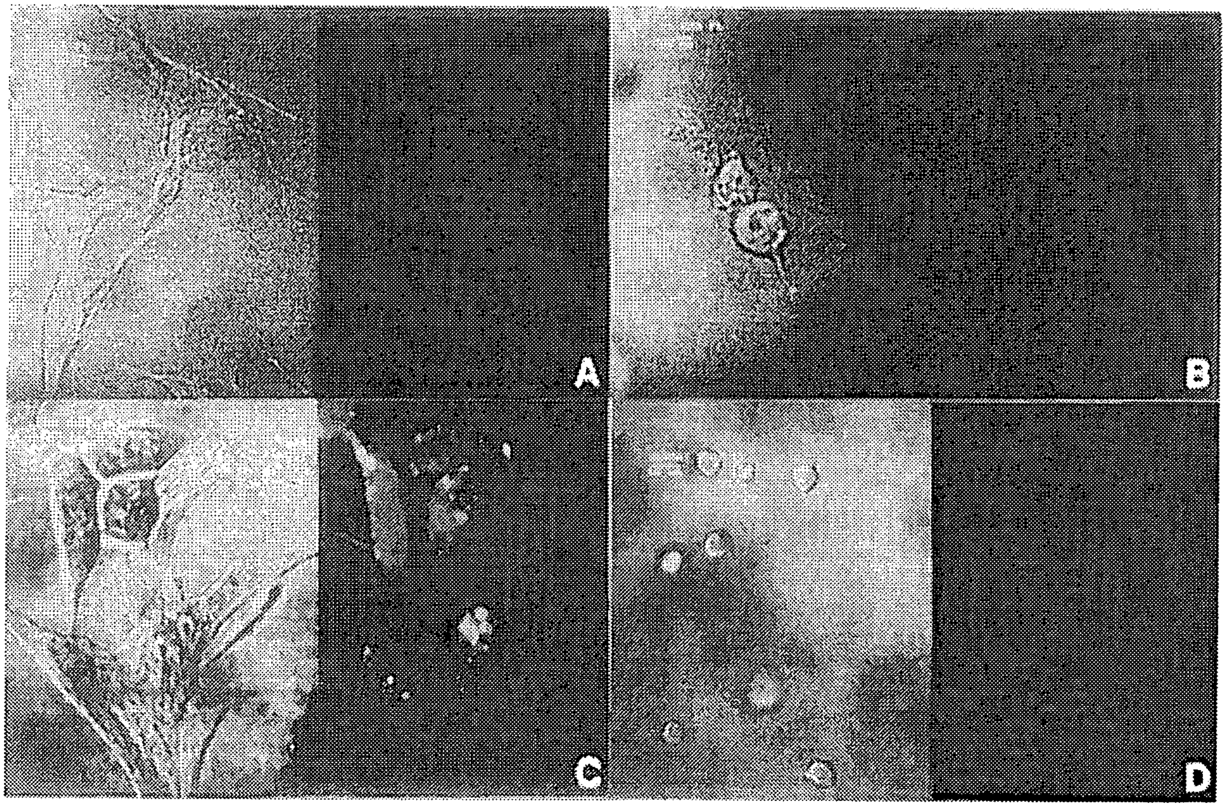

Figure 6. Confocal microscopy of mammalian cells incubated with 1 μM fluorescein derivatives for 4 h at 37 °C: (A) p6 cells incubated with fluorescein alone; (B) p6 cells incubated with fluoresceinated PNA1; (C) p6 cells incubated with fluoresceinated PNAP1 peptide conjugate; (D) Jurkat cells incubated with fluoresceinated PNAP1 peptide conjugate. The right-hand side is the fluorescence image, and the left-hand side is the phase contrast image of the same field.

sequence did not adversely affect the hybridization efficiency of the PNA molecule to its complementary DNA target, in agreement with previous findings (*25, 26*).

Serum Stability of the PNA–Peptide Conjugate. The conjugate was designed to be resistant to both proteases and nucleases, as the PNAs are not degraded by either the proteases or the nucleases (*6*), and D-peptides are also resistant to proteases (*13*). Samples of PNAP1 incubated in serum were extracted at various times and analyzed by HPLC. As hypothesized in the design of PNAP1, the 24 h chromatogram displayed no detectable degradation (Figure 3B).

Cellular Uptake Studies with [^{14}C]PNA and [^{14}C]-PNA–Peptide Conjugates. The rationale behind the construction of the conjugate was that the peptide part would bind to IGF1R, a cell surface receptor which is overexpressed in a variety of oncogenic cell lines, and that would lead to internalization of the molecule via an endocytotic mechanism. Once inside the cells, the PNA may then find its target nucleic acid and hybridize. Therefore, cellular uptake of PNA–peptide conjugates was compared among nontransfected BALB/c 3T3 murine cells, which express modest levels of murine IGF1R, p6 cells, which are BALB/c 3T3 murine cells that overexpress human IGF1R (*14*), and human Jurkat cells, which display very low expression of IGF1R (*17*), as a negative control. BALB/c 3T3 cells, the parental line from which p6 cells were derived, displayed low uptake of [^{14}C]-PNAP1, which was still 4-fold more than BALB/c 3T3 uptake of [^{14}C]PNAP2 and [^{14}C]PNA1 controls (Figure 5A). Uptake measurements with p6 cells showed much greater uptake than observed with BALB/c 3T3 cells, with 5-fold greater uptake of [^{14}C]PNAP1 than of [^{14}C]-PNAP2 or [^{14}C]PNA1 controls (Figure 5B). In contrast, uptake experiments with Jurkat cells, which express virtually no IGF1R, displayed low uptake, with little preferential uptake of [^{14}C]PNAP1 relative to the controls (Figure 5C).

Cellular Uptake Studies with PNA–Fluorescein and PNA–Peptide–Fluorescein Conjugates. When the PNAP1 conjugate and the control PNA1 sequence were fluoresceinated, cellular uptake could be studied by fluorescence under a laser confocal microscope. The fields observed for fluorescence signals were also observed under phase contrast settings. Confocal microscopy allowed observation of different planar sections of the cells, which distinguished between intracellular fluorescence and cell surface fluorescence. This constitutes a significant advantage over conventional fluorescence microscopy, which cannot differentiate between signals inside *vs* outside the cells. Incubation of p6 cells with fluorescein (hydrolyzed FITC) alone (Figure 6A) or with fluoresceinated PNA1 (Figure 6B) showed only very weak cell-associated fluorescence, demonstrating that these molecules were negligibly taken up by p6 cells. This observation implies that uptake was not mediated by the fluorescein moiety. In contrast, incubation of p6 cells with fluoresceinated PNAP1 conjugate showed bright fluorescence signals associated with intracellular structures, including the nuclei (Figure 6C). Since the observed field was a planar section from inside of the cells, the PNAP1 conjugate was presumed to have been internalized. Finally, when Jurkat cells were exposed to the fluoresceinated PNA–peptide conjugate, negligible cell-associated fluorescence signals were observed (Figure 6D), consistent with the hypothesis that the uptake of the PNA–peptide conjugate PNAP1 would be mediated by the JB9 peptide moiety. Figure 7 shows the fluorescence image from the right side of Figure 6C cells superimposed upon the phase contrast image from the left side of Figure 6C, illustrating that the fluorescein signals came from compartments within the cells themselves.

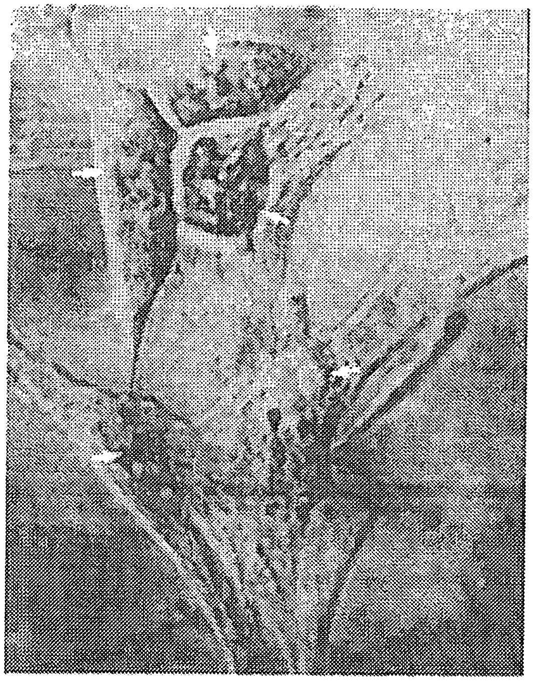

Figure 7. Confocal microscopy of p6 cells incubated with fluoresceinated PNAP1 peptide conjugate. The fluorescence image and the phase contrast image of the same fields shown in Figure 6C are superimposed to illustrate the location of the fluorescence signals within the cellular boundaries. Nuclei are indicated by arrows.

DISCUSSION

The potential of antisense/antigene oligonucleotides as a novel class of therapeutic agents lies in the predictability and specificity of the complementary binding between the intracellular target RNA or DNA and the exogenously applied oligonucleotides by hydrogen bonding. The development of antisense/antigene oligonucleotides as therapeutic agents has focused on chemically modifying the oligonucleotides with the aim of enhancing biological stability and intracellular uptake, without compromising the hybridization efficiency or the binding specificity.

The JB9 peptide moiety, modeled to imitate an IGF1R binding domain of IGF1 (*13*), was hypothesized to enable cell specific targeting of an oligonucleotide, followed by receptor-mediated endocytosis, for the purpose of significantly increasing internalization. Previously, this rationale led us to synthesize a phosphorothioate DNA–JB3 conjugate (*27*). However, PNAs fit the criteria for therapeutic oligonucleotides better than phosphorothioates, with the exception of poor cellular uptake (*9, 10*). To solve this problem, a PNA–IGF1 analog conjugate was synthesized, which resulted in targeted and elevated cellular uptake.

For therapeutic application of oligonucleotide based agents, it is essential to develop an automated synthetic protocol. PNAs are typically synthesized on polystyrene-based solid supports, utilizing Boc protection, similar to what was developed for normal peptide synthesis (*16, 28*). Alternatively, a PNA synthesis has been reported utilizing Fmoc protection (*29*). Manual assembly of a PNA–peptide conjugate on a solid support has been described before, utilizing Boc protection for both segments (*25*).

The peptide segment, containing naturally occurring L-amino acids, was designed to be a substrate for protein kinase A, which phosphorylates the serine residue. Since PNAs have potential for application as antigene/antisense agents against many genetic targets, and PNA–peptide conjugates might overcome the problem of poor cellular uptake, it would be much more efficient to synthesize them automatically.

As a significant step in that direction a semiautomated synthesis protocol has been developed. The synthesis of the conjugate was achieved by integration of Fmoc chemistry to assemble the peptide moiety on a solid support and then Boc chemistry to extend the PNA segment. The other significant improvement was the use of D-amino acids for the peptide section, so that the IGF1 D-analog domain of the conjugate would be as resistant to degradation as the PNA segment, for optimum bioavailability.

The key problem in achieving solid phase synthesis of a biologically stable conjugate was finding the right combinations of protecting groups, so that the two synthetic strategies would be compatible. This task was made more difficult by the inclusion of D-amino acids, as commercially available choices for them are narrow. Utilization of Mob, Bzl, and Z protecting groups was found to be mutually compatible for synthesis of the two segments, resulting in an excellent yield of deprotected, cyclized conjugate. A cyclized peptide domain is essential in the conjugate to limit the flexibility of the JB9 segment, optimizing binding to the cell surface IGF1 receptor. The presence of the cyclized PNA–peptide conjugate was demonstrated by the negative DTNB assay, a homogeneous peak in the analytical HPLC chromatogram, a single peak of predicted molecular mass in the MALDI-TOF mass spectrum, and a single band on a denaturing SDS–polyacrylamide gel, used for the first time to characterize a PNA–peptide conjugate.

Measurements of cellular uptake and compartmentalization demonstrated specificity for the IGF1 analog, and elevated uptake of the PNA–IGF1 analog conjugate was dependent on the level of IGF1R expression. The conjugate was stable in serum over 24 h, which is essential for bioavailability. Furthermore, the peptide adduct did not inhibit hybridization of the PNA segment to complementary DNA. Finally, neither the radiolabeled nor the fluorescent conjugates were toxic to the cell lines tested. This is the first effective demonstration of cell specific targeting and delivery of PNAs.

For oligonucleotides to be applied successfully as antisense/antigene therapeutics, intracellular delivery is essential. Among the available assortment of modified oligonucleotide analogs, PNAs look very promising due to their favorable hybridization characteristics and biological stability. The successful delivery of the PNA–peptide conjugate inside cells without the need for microinjection opens up the opportunity for developing these oligonucleotide analogs as therapeutic agents. The next logical step is to evaluate their antisense efficacy in cell culture. This delivery strategy may be further improved with more potent ligands to achieve better cellular uptake, and other disease-related or tissue-specific cell surface receptors may be targeted in the future to concentrate PNA conjugates into other cell and tissue types.

ACKNOWLEDGMENT

We thank Dr. Michael Kligshteyn for synthesis of DNA analogs, Dr. Rolf Berg for synthesis of peptide nucleic acids, and Dr. Janet Smith for a critical reading of the manuscript. The p6 cell line was kindly supplied by Dr. Christopher Sell, while Dr. Danuta Kozbor provided the Jurkat cell line. This research was supported by National Institutes of Health Grants CA42960 and CA60139 to E.W.

LITERATURE CITED

(1) Wickstrom, E., Ed. (1991) *Prospects for Antisense Nucleic Acid Therapy of Cancer and AIDS*, Wiley-Liss, New York.

(2) Crooke, S. T., and Lebleu, B., Eds. (1993) *Antisense Research and Applications*, CRC Press, Boca Raton, FL.

(3) Agrawal, S., Ed. (1996) *Antisense Therapeutics*, Humana Press, Totowa, NJ.

(4) Nielsen, P. E., Egholm, M., Berg, R. H., and Buchardt, O. (1991) Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide. *Science 254*, 1497–1500.

(5) Egholm, M., Buchardt, O., Christensen, L., Behrens, C., Freier, S. M., Driver, D. A., Berg, R. H., Kim, S. K., Norden, B., and Nielsen, P. E. (1993) PNA hybridizes to complementary oligonucleotides obeying the Watson-Crick hydrogen-bonding rules. *Nature 365*, 566–568.

(6) Demidov, V. V., Potaman, V. N., Frank-Kamenetskii, M. D., Egholm, M., Buchard, O., Sonnichsen, S. H., and Nielsen, P. E. (1994) Stability of peptide nucleic acids in human serum and cellular extracts. *Biochem. Pharmacol. 48*, 1310–1313.

(7) Møllegaard, N. E., Buchardt, O., Egholm, M., and Nielsen, P. E. (1994) Peptide nucleic acid-DNA strand displacement loops as artificial transcription promoters. *Proc. Natl. Acad. Sci. U.S.A. 91*, 3892–3895.

(8) Hanvey, J. C., Peffer, N. J., Bisi, J. E., Thomson, S. A., Cadilla, R., Josey, J. A., Ricca, D. J., Hassman, C. F., Bonham, M. A., Au, K. G., Carter, S. G., Bruckenstein, D. A., Boyd, A. L., Noble, S. A., and Babiss, L. E. (1992) Antisense and antigene properties of peptide nucleic acids. *Science 258*, 1481–1485.

(9) Bonham, M. A., Brown, S., Boyd, A. L., Brown, P. H., Bruckenstein, D. A., Hanvey, J. C., Thomson, S. A., Pipe, A., Hassman, F., and Bisi, J. E. (1995) An assessment of the antisense properties of RNase H-competent and steric-blocking oligomers. *Nucleic Acids Res. 23*, 1197–1203.

(10) Gray, G. D., Basu, S., and Wickstrom, E. (1997) Transformed and immortalized cellular uptake of oligodeoxynucleoside phosphorothioates, 3′ alkylamino oligodeoxynucleotides, 2′-O-methyl oligoribonucleotides, oligodeoxynucleoside methylphosphonates, and peptide nucleic acids. *Biochem. Pharmacol. 48*, 1465–1476.

(11) Pardridge, W. M., Boado, R. J., and Kang, Y-S. (1995) Vector-mediated delivery of a polyamide ("peptide") nucleic acid analogue through the blood-brain barrier *in vivo*. *Proc. Natl. Acad. Sci. U.S.A. 92*, 5592–5596.

(12) Baserga, R. (1995) The insulin-like growth factor 1 receptor: a key to tumor growth? *Cancer Res. 55*, 249–252.

(13) Pietrzkowski, Z., Wernicke, D., Porcu, P., Jameson, B. A., and Baserga, R. (1992) Inhibition of cellular proliferation by peptide analogues of insulin-like growth factor 1. *Cancer Res. 52*, 6447–6451.

(14) Pietrzkowski, Z., Sell, C., Lammers, R., Ullrich, A., and Baserga, R. (1992) Roles of insulinlike growth factor 1 (IGF-1) and the IGF-1 receptor in epidermal growth factor-stimulated growth of 3T3 cells. *Mol. Cell. Biol. 12*, 3883–3888.

(15) Wagner, E., Zenke, M., Cotten, M., Beug, H., and Birnstiel, M. L. (1990) Transferrin-polycation conjugates as carriers of DNA uptake into cells. *Proc. Natl. Acad. Sci. U.S.A. 87*, 3410–3414.

(16) Christensen, L., Fitzpatrick, R., Gildea, B., Petersen, K. H., Hansen, H. F., Koch, T., Egholm, M., Buchardt, O., Nielsen, P. E., Coull, J., and Berg, R. (1995) Solid-phase synthesis of peptide nucleic acids. *J. Pept. Sci. 3*, 175–183.

(17) Lal, R. B., Rudolph, D. L., Folks, T. M., and Hooper, W. C. (1993) Over expression of insulin-like growth factor receptor type-1 in T-cell lines infected with human T-lymphotropic virus types-I and -II. *Leukemia Res. 17*, 31–35.

(18) Ullrich, A., Gray, A., Tam, A. W., Yang-Feng, T., Tsubokawa, M., Collins, C., Henzel, W., Bon Le, T., Katchuria,

S., Chen, E., Jacobs, S., Francke, D., Ramachandran, J., and Fujita-Yamaguchi, Y. (1986) Insulin-like growth factor 1 receptor primary structure: comparison with insulin receptor suggests structural determinations that define functional specificity. *EMBO J. 5*, 2503–2512.

(19) Sarin, V. K., Kent, B. H., Tam, J. P., and Merrifield, R. B. (1981) Quantitative monitoring of solid phase peptide synthesis by the ninhydrin reaction. *Anal. Biochem. 117*, 147–157.

(20) Ellman, G. L. (1959) Tissue sulfhydryl groups. *Arch. Biochem. Biophys. 82*, 70-77.

(21) Pieles, U., Zurcher, W., Schar, M., and Moser, H. E. (1993) Matrix-assisted laser desorption ionization time-of flight mass spectrometry: a powerful tool for the mass and sequence analysis of natural and modified oligonucleotides. *Nucleic Acids Res. 21*, 3191–3196.

(22) Laemmli, U. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature 227*, 680–685.

(23) Hughes, J. A., Avrutskaya, A. V., Brouwer, K. L. R., Wickstrom, E., and Juliano, R. L. (1995) Radiolabeling of methylphosphonate and phosphorothioate oligonucleotides and evaluation of their transport in everted rat jejunum sacs. *Pharm. Res. 12*, 817–824.

(24) Wickstrom, E., Bacon, T. A., and Wickstrom, E. L. (1992) Down-regulation of c-myc antigen expression in lymphocytes of E*μ*-c-*myc* transgenic mice treated with anti-c-*myc* DNA methylphosphonate. *Cancer Res 52*, 6741–6745.

(25) Koch, T., Naesby, M., Wittung, P., Jørgensen, M., Larsson, C., Buchardt, O., Stanley, C. J., Nordén, B., Nielsen, P. E., and Ørum, H. (1995) PNA-peptide chimerae. *Tetrahedron Lett. 36*, 6933–6936.

(26) Soukchareun, S., Tregear, G. W., and Haralambidis, J. (1995) Synthesis and evaluation of nuclear targeting peptide-antisense oligodeoxynucleotide conjugates. *Bioconjugate Chem. 6*, 43–53.

(27) Basu, S., and Wickstrom, E. (1995) Solid phase synthesis of a D-peptide-phosphorothioate oligodeoxynucleotide conjugate from two arms of a polyethylene glycol-polystyrene support. *Tetrahedron Lett. 36*, 4943–4946.

(28) Erickson, B. W., and Merrifield, R. B. (1976) Solid phase peptide synthesis. In *The Proteins*, 3rd ed. (H. Neurath and R. L. Hill, Eds.) pp 255–257, Academic Press, New York.

(29) Thomson, S. A., Josey, J. A., Cadilla, R., Gaul, M. D., Hassman, C. F., Luzzio, M. J., Pipe, A. J., Reed, K. L., Ricca, D. J., Wiethe, R. W., and Noble, S. A. (1995) Fmoc mediated synthesis of peptide nucleic acids. *Tetrahedron 51*, 6179–6194.

BC9700650